Manual of Brazilian Portuguese Linguistics
MRL 21

Manuals of Romance Linguistics

Manuels de linguistique romane
Manuali di linguistica romanza
Manuales de lingüística románica

Edited by
Günter Holtus and Fernando Sánchez-Miret

Volume 21

Manual of Brazilian Portuguese Linguistics

Edited by
Johannes Kabatek and Albert Wall

DE GRUYTER

ISBN 978-3-11-040586-6
e-ISBN (PDF) 978-3-11-040595-8
e-ISBN (EPUB) 978-3-11-040606-1

Library of Congress Control Number: 2022935348

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the Internet at http://dnb.dnb.de.

© 2022 Walter de Gruyter GmbH, Berlin/Boston
Cover image: © Marco2811/fotolia
Typesetting: jürgen ullrich typosatz, Nördlingen
Printing and binding: CPI books GmbH, Leck

www.degruyter.com

Manuals of Romance Linguistics

The international handbook series *Manuals of Romance Linguistics* (*MRL*) offers an extensive, systematic and state-of-the-art overview of linguistic research in the entire field of present-day Romance Studies.

MRL aims to update and expand the contents of the two major reference works available to date: *Lexikon der Romanistischen Linguistik* (*LRL*) (1988–2005, vol. 1–8) and *Romanische Sprachgeschichte* (*RSG*) (2003–2008, vol. 1–3). It will also seek to integrate new research trends as well as topics that have not yet been explored systematically.

Given that a complete revision of *LRL* and *RSG* would not be feasible, at least not in a sensible timeframe, the *MRL* editors have opted for a modular approach that is much more flexible:

The series will include approximately 60 volumes (each comprised of approx. 400–600 pages and 15–30 chapters). Each volume will focus on the most central aspects of its topic in a clear and structured manner. As a series, the volumes will cover the entire field of present-day Romance Linguistics, but they can also be used individually. Given that the work on individual MRL volumes will be nowhere near as time-consuming as that on a major reference work in the style of LRL, it will be much easier to take into account even the most recent trends and developments in linguistic research.

MRL's languages of publication are French, Spanish, Italian, English and, in exceptional cases, Portuguese. Each volume will consistently be written in only one of these languages. In each case, the choice of language will depend on the specific topic. English will be used for topics that are of more general relevance beyond the field of Romance Studies (for example Manual of Language Acquisition or Manual of Romance Languages in the Media).

The focus of each volume will be either (1) on one specific language or (2) on one specific research field. Concerning volumes of the first type, each of the Romance languages – including Romance-based creoles – will be discussed in a separate volume. A particularly strong focus will be placed on the smaller languages (linguae minores) that other reference works have not treated extensively. MRL will comprise volumes on Friulian, Corsican, Galician, among others, as well as a Manual of Judaeo-Romance Linguistics and Philology. Volumes of the second type will be devoted to the systematic presentation of all traditional and new fields of Romance Linguistics, with the research methods of Romance Linguistics being discussed in a separate volume. Dynamic new research fields and trends will yet again be of particular interest, because although they have become increasingly important in both research and teaching, older reference works have not dealt with them at all or touched upon them only tangentially. MRL will feature volumes dedicated to research fields such as Grammatical Interfaces, Youth Language Research, Urban Varieties, Computational Linguistics, Neurolinguistics, Sign Languages or Forensic Linguistics. Each volume will offer a

structured and informative, easy-to-read overview of the history of research as well as of recent research trends.

We are delighted that internationally-renowned colleagues from a variety of Romance-speaking countries and beyond have agreed to collaborate on this series and take on the editorship of individual MRL volumes. Thanks to the expertise of the volume editors responsible for the concept and structure of their volumes, as well as for the selection of suitable authors, MRL will not only summarize the current state of knowledge in Romance Linguistics, but will also present much new information and recent research results.

As a whole, the MRL series will present a panorama of the discipline that is both extensive and up-to-date, providing interesting and relevant information and useful orientation for every reader, with detailed coverage of specific topics as well as general overviews of present-day Romance Linguistics. We believe that the series will offer a fresh, innovative approach, suited to adequately map the constant advancement of our discipline.

Günter Holtus (Lohra/Göttingen)
Fernando Sánchez-Miret (Salamanca)
July 2022

Acknowledgements

This, the first handbook of Brazilian Portuguese Linguistics written entirely in English for a wide, international readership, has been challenging but immensely satisfying to coordinate. It is beyond the scope of this preliminary note to dwell on the many joys, and occasional sorrows, that we have experienced throughout the years of its making, and it goes without saying that it would not have been possible without the unceasing support of many individuals who shared our conviction that such a handbook constitutes an important and significant addition to this field of scholarship. First and foremost, we would like to express our profoundest gratitude for the substantial contributions made by José da Silva Simões during the initial phases of the project. José was a member of the editorial team and his ideas were decisive not only in shaping the volume but also in recruiting specialists for many of the chapters, as well as in the considerable energy he invested in revising the early versions of many chapters. Although other obligations took him away from the project at a later stage, José's enthusiasm and intellectual vigor has left its stamp on each of these pages. Muito obrigado por tudo, José.

We also want to express our sincerest gratitude to all the wonderful authors herein for accepting our invitation to participate, taking our proposals for their chapters into consideration, bringing in their own ideas, and sharing their ample expertise and personal insights into their respective areas of research. It was especially rewarding to meet many of you and to discuss the ongoing work of this project personally in August 2019 at a workshop in Brasilia dedicated to the preparation of the handbook; indeed, this was crucial in giving the volume its final structure and orientation. Yoselin Henriques Pestana, who also collaborated in the formatting process, and María Luisa Gago Iglesias from the University of Zurich dedicated a great deal of time and effort in organizing this remarkable event. Thank you! We would also like to thank the Universities of Zurich (which among many other things funded the Brasilia workshop) and the University of Vienna for supporting our work.

John Barlow was responsible for proofreading of the entire book and we are very grateful for his experienced and thorough engagement throughout.

Finally, we would like to thank the editors of the MRL series, Günter Holtus and Fernando Sánchez-Miret, for their expertise and regular feedback, as well as the team at de Gruyter for their support and patience in challenging moments. Ulrike Krauß and Gabrielle Cornefert kept a keen eye on all those important details that might otherwise have escaped our attention and were always available for our questions, Meiken Endruweit proofread the whole manuscript scrupulously, and Sabina Dabrowski steered us marvelously through the final production phase.

Individual acknowledgments of authors, by chapter

Chapter and author(s)	Thanks to…
Chapter 1 Cristina Altman Ataliba T. de Castilho	Albert Wall, for significant contributions.
Chapter 6 Mário Viaro	Michael Jones Ferreira, for reviewing and comments.
Chapter 8 Charlotte Galves Marilza de Oliveira	Karina Molsing, for proofreading and CNPq (grant 313531/2018–8) for financial support.
Chapter 12 Dermeval da Hora Elisa Battisti	Caio Cesar Martino, for translation services and Volker Noll for reviewing and comments.
Chapter 15 Maria da Graça Krieger	CNPq, for financial support.
Chapter 18 Patricia Carvalhinhos	Tathiana Ferguson Motheo for translation services, Amelina Pedrali de Aquino for proofreading and Adriana Tavares Lima and Maria da Graça Krieger for reviewing and comments.
Chapter 19 José Luiz Fiorin	Vanderci Aguilera, Cristina Altman and Ronald Beline, for reviewing and comments, Karina Molsing for translation.

Table of Contents

Albert Wall and Johannes Kabatek
0 Introduction: the state of the art in Brazilian Portuguese linguistics —— 1

Cristina Altman and Ataliba T. de Castilho
1 Brazilian Portuguese linguistics: an overview —— 23

Tânia Lobo
2 The social history of Brazilian Portuguese —— 53

Dante Lucchesi and Alan Baxter
3 The history of linguistic contact underlying Brazilian Portuguese —— 85

Volker Noll
4 Historical phonetics and phonology —— 113

Mário Eduardo Viaro
5 Historical morphology —— 133

Charlotte Galves and Célia Regina dos Santos Lopes
6 Historical syntax —— 155

Américo Venâncio Lopes Machado Filho
7 The history of the lexicon —— 193

Charlotte Galves and Marilza de Oliveira
8 The debate on Brazilian and European Portuguese —— 219

Vanderci de Andrade Aguilera
9 Dialectology and linguistic geography —— 251

Uli Reich and Ronald Beline Mendes
10 Sociolinguistics —— 283

John M. Lipski
11 Brazilian Portuguese: contemporary language contacts —— 309

Dermeval da Hora and Elisa Battisti
12 Sound-related aspects of Brazilian Portuguese —— 337

Carlos Alexandre Gonçalves
13 Morphology — 369

Sanderléia Roberta Longhin and Maria Clara Paixão de Sousa
14 Syntax — 401

Maria da Graça Krieger
15 Lexicon — 435

Rodolfo Ilari and Renato Miguel Basso
16 On the history of semantic studies in Brazil — 459

Alessandra Castilho da Costa
17 Spoken *vs.* written language — 487

Patricia Carvalhinhos
18 Onomastics and toponomastics — 513

José Luiz Fiorin
19 Linguistic policy and the orthographic agreement — 553

Elisângela N. Teixeira and Michele C. dos S. Alves
20 Psycholinguistic studies: language acquisition and processing — 579

Index — 615

Albert Wall and Johannes Kabatek

0 Introduction: the state of the art in Brazilian Portuguese linguistics

1 Brazilian Portuguese linguistics as a subject on its own

This is the first comprehensive manual on Brazilian Portuguese (BP) linguistics in English and the second manual on Portuguese (following the publication of the *Manual de Linguística Portuguesa* in 2016) in the De Gruyter MRL series.[1] This perhaps requires some explanation: why another manual about the same language? There are at least two reasons for this. The first is the language itself, and the fact that BP is not just a colonial continuation of Portuguese with a few particular characteristics, but rather —paradoxically— both a different language and the same language. The differences are in part quite old, but many of them have changed their status recently and have thus become visible, to be seen not only in substandard varieties. The second reason, closely related to the first, is the current state of the art in BP linguistics. In recent times, linguists from across the world have shown a particular interest in BP, and in Brazil an enormous body of work on a wide range of linguistic areas has shed light on a reality which was quite unknown in the past. Much of what has been studied by linguists in Brazil has remained to a great extent hidden from a wider international public, and hence one of the aims of this book is to offer insights into recent research not only on BP linguistics in general but also work on BP produced in Brazil. Furthermore, over the last two decades, two large-scale projects have achieved remarkable milestones in publishing comprehensive and condensed results in two multi-volume series: in the first decade of the 21st century, a comprehensive "Grammar of spoken Brazilian Portuguese" was published, and in the following decade, the "History of Brazilian Portuguese", see ↗1 Brazilian Portuguese linguistics: an overview for full references and project descriptions.

As for the language itself, the paradox noted above is a way of avoiding conflicts as well as misunderstandings. Emotional and ideological conflicts arise when linguists claim that typologically BP is so different from European Portuguese (EP) that the common name is somehow misleading: even if they are historically linked and still very closely related, EP and BP are at the same time two different linguistic reali-

1 Despite the title *Manual de Linguística Portuguesa*, the editors explain in the introduction that, given the publication of a second manual, the object of the first is European Portuguese with occasional references to Brazilian Portuguese (Martins 2016, 1).

https://doi.org/10.1515/9783110405958-001

ties. Why does such a simple and objective statement produce emotional reactions? Basically, because it leads to a misunderstanding between a judgement about linguistic difference and what is understood as an implicit political opinion about possible consequences, something like an aggression against the maintenance of unity. But the maintenance of unity on an abstract level, as well as the maintenance of a common orthographic system, are political issues of education and cultural organization and, one might argue, with objective differences or similarities or with historical links in favor of or against such maintenance. However, the decision will not emerge as an objective, causal consequence of these facts. Issues like the difference between EP and BP, as well as the political debate about Lusophone unity or diversity, will be addressed in this manual alongside descriptive presentations of linguistics facts, but both views will be separated as far as possible.

In debates in general linguistics as well as in Romance or in Portuguese linguistics, BP normally appears as a reality of its own. Instead of mentioning a few particularities of BP in contrast to EP, BP can be described as an autonomous diasystem that emerged historically from EP but which presents so many different features that an independent description seems to be more efficient, at least as a first step. Of course, comparison with EP makes sense on all levels of linguistic description, but it should always be a comparison of two realities on the same level, not between a default case and putative or claimed deviation.

BP is a Romance language that is part of the "New Romania", a consequence of the Portuguese colonization of South America. Its base is 16th century Portuguese, modified by internal evolutions as well as by contacts with the native peoples and with enslaved people deported from Africa, plus contact with other European languages present in Brazil, especially since the 19th century. The contacts might have had limited consequences, and these mainly in the lexicon, and indeed, descriptions of the particularities of BP until the 19th century and of historical documents might lead us to think so, yet a closer look shows that something new has emerged beneath the surface of what appears in writing. By contrast to the Spanish crown, which created relatively independent infrastructures with universities, book printing and important colonial urban centers in the Americas, Portugal allowed neither printing nor the creation of a university in Brazil: the colony was strictly linked to Lisbon as the political and economic center, and to the intellectual primacy of the university of Coimbra, the only place where Brazilian people could pursue university studies. As early as the 19th century, some of the fundamental innovations of BP seem to have been achieved in the spoken, substandard language, in the first place the loss of many properties of Romance pro-drop languages such as an ongoing obligatorization of overt subjects and the rise of *você* and *a gente* as subject pronouns. As in Middle French, this favors the loss of endings and entails several further evolutions that continue to the present day. In the 20th century, BP exists with a completely different prosody from its European counterpart, a fact that is due not only to changes in the rhythmic structure of BP, but also in EP with its ongoing reduction of unstressed vowels, especially in the variety of Lisbon.

In Brazil, several external developments have enhanced processes of prestige changes of vernacular features during the second half of the 20th century. While Rio de Janeiro, the traditional capital, used to function as a bridge to Europe, even after Brazilian independence, the artificial creation of the new capital Brasilia in the 1950s did not create a new center of gravity but rather produced a "regionalization" of Rio's *norma carioca*, giving new weight to São Paulo, the largest urban center of the country, where mass media and economic power concentrate.

BP in the 21st century is still in a process of implicit and explicit debates about how to reorder its diasystem. Tensions between the colonial past, reflected in formal and written language, and the oral vernacular are nowadays above all tensions between socioeconomic classes within the country, and the answer to the question of what BP really is will always depend on the specific referent within the architecture of its varieties. Even among linguists, a normative view is not always strictly distinguished from descriptive approaches. However, we currently have at our disposal a large number of studies that allow us to recognize the multiple phenomena that characterize BP. Even if BP is spoken across half a continent, traditional linguistics stated that the vertical, diastratic variation was more clearly marked than the horizontal, diatopic one, as can frequently be observed in colonial situations. But this is also an impression due to the tradition of looking mainly at urban varieties in dialectological studies. More recent approaches also show significant regional variation. At the same time, urban BP is far from unitary, and Brazilian cities are complex social and linguistic conglomerates. Contacts—historical and actual—are further factors in the diversity of the present configuration of BP. This handbook aims to provide comprehensive information about this very broad panorama.

2 An overview

The moment seems perfect for offering an overview of BP. Recent decades have seen an enormous number of studies on BP, both within and outside Brazil. In Romance and in general linguistics, phenomena that can be observed in Brazil challenge linguistics. Romance linguists are impressed by the dynamic current linguistic "laboratory" of Brazil, offering insights into changes in progress similar to ones that in other cases remain accessible only indirectly through the observation of their historical consequences. Grammaticalization theory is confronted with evolutions that seem to contradict currently accepted assumptions. Formal semanticists look with astonishment at BP data, and syntacticians are fascinated by BP structures. All this has become possible due to a great deal of work involving data collection and description over the last few decades, and due to several collaborative research projects yielding an immense body of findings: from an ongoing project on variation in educated urban language (NURC) which began in the 1960s, several groups of researchers have contributed to a much better and much more systematic knowledge as to what BP really is. The pro-

ject on the 'Grammar of spoken (Brazilian) Portuguese' (*Gramática do Português Falado*), the project 'Towards the History of Brazilian Portuguese' (*Para a História do Português Brasileiro*, PHPB), as well as other exhaustive projects such as one on the language of São Paulo (see ↗1 Brazilian Portuguese linguistics: an overview) have generated large amounts of corpora and an impressive quantity of studies. What is absolutely remarkable in these projects is that they are collaborative in a double sense: the example of the PHPB project shows that in Brazil something that in many places would be an unthinkable enterprise was possible: the collaboration not only across disciplines, involving historians, sociolinguists and systemic linguists, but also collaboration within the discipline, when functional, structural and generative linguists have come together towards an attempt to offer different and mutually enriching views on the same phenomena.

Unfortunately, a lot of this production is not easily accessible on the global market due to the scarce international distribution of print publications and due to the fact that most of the studies were published in Portuguese, a language of course familiar to specialists but not even particularly widely known among Romance linguists, let alone general linguists.

2.1 Corpora

A well-founded description of any linguistic reality cannot be achieved without a solid empirical basis. Important building blocks in any such basis are corpora that reflect linguistic production in a given language. Many of the facts that specialists of BP these days take for granted would not have been known without the compilation, information processing and analysis of corpora of different sizes and composition, be it with regard to the diachrony of the language, with regard to its variation at different points in time, or the recurrent patterns and forms that speakers use. New corpora are constantly being created, existing corpora are updated and refined continuously, and technological innovations are changing the overall picture rapidly. The goal of this section is thus to serve as a starting point for the interested reader and to describe some of the most important current activities and resources available, rather than to give a full account of BP corpus linguistics. Nevertheless, we provide some details below, given that this volume does not include a chapter dedicated to corpus linguistics, although many chapters do make reference to these corpora, without introducing them in great detail. An overview of (European) Portuguese corpus linguistics can be found, for instance, in the handbook article Mendes (2016) or in the collective volume Berber Sardinha/São Bento Ferreira (2014). Most of the resources described below were produced in Brazil and reflect a small fraction of BP text production. However, we also included several interesting resources produced abroad that contain significant amounts of BP data, and also one very notable corpus that, despite being compiled in Brazil, represents historical EP.

Currently, the two largest and best (automatically) preprocessed databases for BP are the Corpus Brasileiro, a one-billion word web-based corpus of Brazilian texts (Berber Sardinha 2010), and the Corpus do Português (Davies/Ferreira 2006–; Davies 2016–).[2] The Corpus Brasileiro is composed entirely of Brazilian texts from a variety of genres automatically retrieved from the internet between 2008 and 2010 and can be accessed via the commercial Sketch Engine platform and explored with the many tools provided there. Another way of accessing it free of charge is through the platform linguateca.pt,[3] where more detailed information on the text types and the composition of the corpus can be found. The Corpus do Português, as the name suggests, also includes other varieties of Portuguese, and has several subparts. The original, diachronic part with texts from the 12th century onwards (texts from Brazil from the 16th century onwards) has about 45 million words, containing EP and BP in roughly equal parts. There are also two larger web-based sub-corpora with over 1 billion words, in each of which the Brazilian part is the largest subpart. The "Web/Dialect" sub-corpus contains over 600 million words, whereas the "News on the Web" sub-corpus is currently of a similar total size, although every month some 35 million words of new texts are added. Both databases, the Corpus Brasileiro and the Corpus do Português are POS tagged and have sophisticated search interfaces. Besides these large-scale, automatically processed corpora, there are also several corpora of spontaneous spoken language that have been recorded and transcribed manually, as well as corpora of older, non-literary texts that have been and are being edited into a diachronic corpus of BP.

As for spoken language, the C-ORAL-BRASIL corpus (Raso/Mello 2012), the Brazilian contribution to the C-ORAL-ROM project (Cresti/Moneglia 2005), is composed of an already-published sub-corpus of informal speech and a sub-corpus of formal speech, which is currently in preparation. Its total size is 500,000 words (45 hours of recordings), which is larger than other corpora of the C-ORAL family. Nevertheless, since the composition and processing of the corpus follows the C-ORAL-ROM standards and criteria, it can be used for cross-linguistic comparison, one of the main goals of this corpus family. The majority of the speakers that have been recorded for this corpus come from the state of Minas Gerais. The database contains conversations as well as dialogues and monologues. The corpus has been automatically tagged for POS and syntactic functions using the PALAVRAS parser, which was adapted to the C-ORAL-BRASIL speech data. The entire corpus including audio files is freely available for download.[4]

Much larger than C-ORAL-BRASIL and much more diversified are the corpora with sociolinguistic profiles. By far the largest of these is the one that emerged from *Norma*

2 https://www.corpusdoportugues.org/, last accessed 11.02.2022.

3 https://www.linguateca.pt/acesso/corpus.php?corpus=CBRAS, last accessed 11.02.2022.

4 http://www.c-oral-brasil.org/, last accessed 11.02.2022.

Urbana Culta (= NURC), a project that investigated the spoken language of speakers with higher education in five state capitals (see ↗1 Brazilian Portuguese linguistics: an overview). The recordings of this project comprise more than 1,500 hours of speech. They were made in the 1970s, and have been transcribed and further processed to different degrees, depending on the kind of analysis used. The corpus includes formal speech, semi-directed interviews and dialogues between informants, which were balanced for gender and age. Being organized in a decentralized way, and probably also because of the sheer size of the corpus, it has never been integrated into a single database. However, large parts of the NURC data are freely available for download on the websites of various local projects.[5] Most of these materials are semi-orthographic transcriptions without further processing, accompanied by the digitalized audio-files. In some cities, some informants from the original project were re-contacted and the interviews were repeated in the 1990s. Furthermore, new speakers were added in order to allow for quantitative diachronic observations. Parallel to the compilation of the NURC corpus, several projects also gathered data on speakers with lower levels of education (see Silva 2015 for a characterization of many of them). For instance, in Rio de Janeiro, 140 interviews were conducted, recorded and transcribed with 20 participants of the *Movimento Brasileiro de Alfabetização* (MOBRAL, 'Brazilian movement for alphabetization'), see Lemle/Naro (1977), hence speakers with no formal education at all. The so-called CENSO/PEUL corpus (named after the projects that compiled it) includes recorded interviews and transcriptions from the 1980s and 2000s, with speakers that had gone through some basic schooling. The core of the corpus is of recorded and transcribed interviews with 64 speakers, including children aged 7–14 (Silva 2015, 171). The MOBRAL and CENSO/PEUL corpora are currently not available on-line. While these and other sociolinguistic corpora have been explored and the analyses published in many theses and articles, and thus have served as crucial contributions to our knowledge and understanding of the linguistic reality of Brazil, the transcriptions and annotations have never been brought together in a unified and searchable digital database. For the NURC corpus, size alone might explain this state of affairs. However, according to Dinah Callou (p.c.), the transcriptions provided in the NURC corpus were in fact never intended to replace the recordings for analysis. It was taken for granted by the researchers that they served merely as a point of reference for the content of the recordings, and that analysis of spoken language should only be done on the original data.

As noted above, this is not the place for a complete account of BP corpus linguistics, not least because such an undertaking would include the entire discipline of computational linguistics, which is beyond the scope of this introduction (but see below). As representative of the many specific initiatives that are currently developing

5 See, for instance, https://fale.ufal.br/projeto/nurcdigital/, https://nurcrj.letras.ufrj.br/, http://nurc.fflch.usp.br/, last accessed 11.02.2022.

corpus data and tools, we might mention LexPorBr (Estivalet 2015), which is interesting because it serves a very different purpose from the ones mentioned so far. It is a "Psycholinguistic Corpus", in the sense that it provides information about BP words which is relevant for psycholinguistic studies and the construction of psycholinguistic experimental stimulus materials. It is based on the NILC/São Carlos corpus (34 million words, journalistic and didactic texts, letters and texts written by pupils in educational environments, Nunes et al. 1996).[6] The query interface of LexPorBr provides different frequency and word structure measures that are relevant for language processing and also provides a tool for systematically creating pseudowords for experimental purposes according to a number of parameters that can be adjusted.

While the NURC and CENSO/PEUL corpora at some point incorporated samples from different decades for diachronic analysis, they are limited to texts from the past 50 years. In order to reveal and document BP use and evolution prior to that, two large-scale projects have built diachronic corpora of BP. These corpora feature different texts types, including, for instance, personal letters, advertisements and theatre plays, which are all less subject to normative correction and hence more likely to contain elements of the spoken language from the past or at least can tell us something about linguistic variation in those times. Chronologically, the first diachronic corpus was built by the project *Para a História do Português Brasileiro* (PHPB 'Towards a History of Brazilian Portuguese'), starting in 1997. The current state of this corpus, plus detailed information about it, can be found in Barbosa (2019), including an appendix of 15 pages with a list of all materials published until 2010, when the second phase of the project was concluded. According to this list, some of the materials are available on-line, some upon request. There is also a corpus platform[7] which contains parts of the materials for download in .pdf or .doc format. The corpus contains texts from the beginning of the 18th until the end of the 20th centuries. During the aforementioned "second phase", the so-called "minimal common corpus" has largely been completed. The minimal corpus has been compiled by local groups at 12 universities throughout the country and contains texts for the entire time span of the corpus, for each of the following text types: personal letters, official letters and private administrative letters (all manuscript sources), and letters to (and from) the editor, advertisements and private administrative letters. It also contains a contrastive corpus of texts, such as letters written by Portuguese authors, theater plays, testaments, and many others (an open list, in principle). The third phase of the project, which started in 2011, is dedicated to the compilation and publication of a complementary corpus, including non-literary printed texts, official documents, and further discourse traditions in line with the research goals of individual local projects. According to Barbosa (2019, 18), the

6 LexPorBr: www.lexicodoportugues.com/, NILC: https://www.linguateca.pt/acesso/corpus.php?corpus=SAOCARLOS, last accessed 11.02.2022.

7 https://sites.google.com/site/corporaphpb/, last accessed 11.02.2022.

PHPB corpus is, in principle, an open-ended endeavor, limited only by the needs of researchers of historical linguistics to have diverse and socio-historically controlled materials at their disposal. Similar to procedures in the sociolinguistic projects, corpus compilation and exploration occurred in parallel in this project. This resulted in a situation in which, whereas the corpus is still "in the making", many studies based on it have already appeared, including a book series synthesizing the information entitled *História do português brasileiro* ('History of Brazilian Portuguese', 10 volumes published, 2 volumes currently in preparation, see ↗1 Brazilian Portuguese linguistics: an overview for details). Thus, although both are corpora of BP, the material of interest for the PHPB corpus is very different from that of the Corpus Brasileiro, for instance. This is basically due to the fact that the PHPB is a "philological" corpus based on laborious transcription work, whereas in the Corpus Brasileiro the texts were collected automatically from the web and invite mainly quantitative studies. The PHPB corpus has had a much more "distributed" and somewhat "preliminary" existence, and awaits further processing by computational linguists.

Coming slightly later than the PHPB project, but overlapping with it in terms of time and participants, the project *Para a História do Português Paulista* (PHPP, 'Towards the History of the Portuguese of São Paulo') comprises 13 local sub-projects from the state of São Paulo which organized themselves into a larger, collaborative project. It is currently building a diachronic corpus for the state of São Paulo, partly re-using materials from PHPB, NURC and other sources, but crucially also doing meticulous philological work and editing new historical texts from different genres. The goal of the corpus section of the project is to have a balanced corpus in terms of text types across eight timespans of 50 years each, from the 17th to the 20th centuries (Simões 2012; Kewitz/Simões 2019). The current state of the corpus can be seen on its webpage, where the materials are organized into a time-text type grid that visualizes this goal in the shape of a table. The already-available materials can be downloaded as text files (word or pdf).[8] Currently, about one third of the cells of the table are filled with materials, most of the documents being various types of letters, announcements and testaments. These documents vary in length from several to several hundred pages. While they all follow the same editing norms, the documents made available on the project homepage sometimes only contain the transcribed texts; in other cases, however, they contain the entire publications from which the texts are taken. As the name of the project and the description of the corpus here suggest, the PHPP follows the spirit of the PHPB corpus, and indeed goes a step further, in that it also aims to make possible the exploration of the 17th century, a very ambitious goal for a corpus of this nature.

Besides the PHPB and PHPP, another significant resource for the investigation of the history of the Portuguese language has been built in Brazil, at the University of

8 http://phpp.fflch.usp.br/corpus, last accessed 11.02.2022.

Campinas: the Tycho Brahe corpus (Galves/Andrade/Faria 2017).[9] The corpus contains mainly texts from EP from the 14th to the 19th centuries, balanced by quantity and text type, and also some texts from Brazil from the 19th century. Its total size is about 3,3 million words. Concerning technical aspects and issues related to annotation, Tycho Brahe is currently the most highly-developed diachronic corpus of Portuguese. The entire corpus is POS tagged and more than one third has been syntactically annotated by hand. Furthermore, the Tycho Brahe platform provides search interfaces for the corpus and many tools and information for diachronic corpus analysis. Its use is free after registration. Besides these larger collaborative initiatives of corpus construction, there are also many individual projects that are currently collecting and processing historical texts, such as the *Grupo de Morfologia Histórica do Português* (GMHP, 'Portuguese Historical Morphology Group') at São Paulo University, which also makes part of its corpora available online.[10]

Historical corpora of Portuguese have also been complied outside Brazil, often with more specific research goals. The Colonia Corpus of Historical Portuguese (Zampieri/Becker 2013), for instance, contains roughly 5 million words (16th–19th century), is balanced in size for EP and BP, and was built for the quantitative comparison of the evolution of EP and BP. The corpus is freely available,[11] POS tagged, and has a query interface. Another example is PorThea, a collection of PE and PB theatre plays from the 18th to the 21st centuries (2,6 million words BP and 700,000 words EP, Rosemeyer 2018). Theater plays have proven to be an interesting data source for the observation of phenomena of spoken language and conversation from those earlier periods for which we have no recordings, and have been explored widely in the study of the morpho-syntactic evolution of BP (Duarte 2012).

2.2 Some leading topics in BP studies

As noted in the introductory section, BP has quite a distinctive status given that from one perspective it is still Portuguese yet at the same time is quite different from EP, having developed its own, absolutely independent linguistic reality. This is also reflected in the fact that while comparison between EP and BP is still an important domain of investigation, the study of BP in its own right has become the default in BP linguistics. Extensive research on all aspects of BP structure has led to many contrasting findings with respect to EP, and these have been taken by many researchers as indicative of a fundamental difference in core grammatical aspects between the two varieties (see, however, ↗8 The debate on Brazilian and European Portuguese for the

9 https://www.tycho.iel.unicamp.br/home, last accessed 11.02.2022.
10 http://www.usp.br/gmhp/Corp.html, last accessed 11.02.2022.
11 https://www.kaggle.com/rtatman/colonia-corpus-of-historical-portuguese, last accessed 28.02.2022.

polemic about EP vs. BP and ↗19 Linguistic policy and the orthographic agreement for linguistic policies for the entire Lusophone world).

The Portuguese spoken in Brazil has been known for its lexical adaptation to the "new world" ever since it arrived at Brazilian shores. Despite the lack of a comprehensive account of the structure and evolution of the BP lexicon or of the contrasts with EP, modern dictionaries document thousands of contrastive lexical entries (Noll 2008, 94–110). A summary and discussion of the first loanwords from Amerindian languages can be found in Noll (2008, 136–152), who also describes important advances in contemporary BP lexicography and points out that the period before the 19th century in particular demands further study (2008, 255–257).

It is a commonplace to observe that a person, on being confronted for the first time with the pronunciation of EP and BP, will be very surprised by the radical differences that she hears. This perception is mainly due to phonetic and phonological processes and prosody, since the segmental inventory of EP and BP can be considered as practically identical (Massini-Cagliari/Cagliari/Redenbarger 2016, 56). As far as processes and prosody are concerned, however, BP and EP follow quite different evolutions, to the point that these processes may affect other parts of grammar and hence stabilize or accelerate ongoing change in different directions.[12] Similar observations have been made with regard to syllable structure, which can be analyzed "at some shallow level of abstraction, if not identical, at least very similar" (Collischonn/Wetzels 2016, 86), where again a series of contrary processes in both varieties produces widely different outcomes—the tendency towards CV-structure in BP and highly complex consonant clusters in EP, to give just one example. Investigation into different outcomes at the prosodic level have led to controversial debate, for instance, with respect to the discussion about rhythm and the distinction between syllable-timed and stress-timed languages. A series of empirical investigations here has shown that both EP and BP are of a mixed type and that rhythm also changes at the dialectal level (Frota/Vigário 2001; Cruz/Frota 2014). However, while EP and BP are not prototypical cases of the stress vs. syllable timing typology, their positions within this typology are quite distant from one another—and in the case of BP also quite distant from the other Romance languages. Exploring such processes a little further, the fact that BP produces many more open syllables than EP means that consequently there are far more putative contexts for vocalic contraction and assimilation at the word boundary. This, in turn, is of relevance for several monosegmental vocalic functional elements, such as definite articles, object clitics and the preposition *a*. For all these elements, it has been claimed that frequent assimilation has the consequences of there being differences in their status in the linguistic systems of BP and EP (Kewitz et al. 2018, 346–352; Wall 2017, 105–152, among others).

12 Of course, there is also considerable variation within EP and BP in this domain.

In the domain of morphology, the partial loss of verb-subject agreement and also strong variation in plural agreement in nominal phrases has developed into a central topic of investigation, especially in historical and sociolinguistic studies, but also beyond (Lemle/Naro 1977; Paixão de Souza 2010; Castilho et al. 2019). It has been investigated on the one hand with respect to the reduction of the verbal paradigm in BP, but also in terms of the syntactic changes in the pronominal domain (see below). Another important research area is the generalization and further grammaticalization of nominal forms that have been introduced into the pronominal system over recent centuries, such as *você* and *a gente*, the loss of certain clitic objects and the replacement of some of these by stressed forms (Lopes 2018). Many other topics of investigation have contributed to revealing the distinct shape of BP grammar in this area, one of which, for instance, is the maintenance of the gerund in verbal periphrases and further evolutions of this construction which in EP has been supplanted by an infinitival construction (Simões 2007).

In the domain of syntax, one of the central topics has been the reorganization of the pronominal system, its sociolinguistic variation, and its formal analysis. Here, the "loss" of the subject pro-drop property, variation in the placement of clitics, and higher flexibility of null objects have constituted an important domain of study, giving BP a unique position within the Romance languages, but also beyond (Cyrino/Nunes/Pagotto 2009; Martins 2018). In the nominal domain, so-called bare nominals have received a lot of attention in recent years, again because BP grammar behaves quite differently from related languages. Being a full-fledged article language, it nevertheless poses much fewer restrictions, if any, on bare subject and object NPs than any other known language of this type (Kabatek 2007; Wall 2017). In both cases, regarding the expression of subject pronouns and bare nominals, BP data have contributed crucially to the revision and refinement of syntactic and semantic theories based on the idea of parameterized universal grammar. Argument structure has also been explored in different directions, for instance, new evolutions in ditransitive sentences and dative alternation (Calindro 2020; Cépeda/Cyrino 2020). Another domain which has seen considerable research activity is that of the replacement of *haver* by *ter* as the default existential verb, with major consequences for the structure of these sentences, and the decrease in the use of reflexive constructions (Avelar 2009; 2018). Intertwined with these phenomena is the broader domain of impersonal sentences, for which detailed studies have once more shown that BP has developed its own grammar. Further central topics include the syntax of negation and the typology of cleft sentences with a wealth of particular configurations (Teixeira de Souza 2015; Modesto 2001; among others). In both cases, BP syntax shows complex patterns which are not found in other Romance languages. These and many other studies have shown time after time that many syntactic generalizations and mechanisms developed on the basis of other Romance and Germanic languages fail to account for the BP data, contributing in this way not only to the study of BP grammar but also to Romance and general linguistics.

Some of the aforementioned syntactic configurations, as well as further related ones, have also constituted a central topic in the study of information structure. This area of research has drawn crucial insights from observations in the work of Pontes (1987), leading to the hypothesis that BP might be developing from a subject-prominent into a subject and topic-prominent language, in the terminology of Li/Thompson (1976), with major consequences for syntactic analysis. This idea has been explored in many studies, such as Vasco (2006), an empirical study on spoken language data from Rio de Janeiro with rich exemplification and variationist analysis of the relevant constructions, and in Silva/Fonseca (2018), an experimental approach which also considers prosody and perception. The results of these studies do not strongly support the somewhat overstated claim in Pontes (1987). Nevertheless, they have made major contributions to the understanding to the configurations first described by Pontes. Further criticism of the underlying idea of a typological contrast to EP has been proposed by Costa (2011), arguing that several of the highlighted constructions also occur in spoken EP, whilst acknowledging that BP also has some quite unique patterns.

A pattern seems to emerge from all the aforementioned domains of investigation, namely, that at first, BP is often perceived as something highly exotic, leading to analyses that set it far apart from the other Romance languages, or which propose radically new analyses for its structures, sometimes accompanied by controversial debate. In a second step, this initial "moment of shock" leads to more in-depth investigations, resulting in more nuanced analyses. Among these radical claims we find the suggestions that BP had a creole basis in the past, that it is developing towards stress-timing, that it is losing the pro-drop property entirely, that it strongly contradicts the Nominal Mapping Parameter, and that it might become a subject and topic-prominent language. A discussion about exotification and eurocentrism (and how to deal with them) would go beyond the scope of this introduction. Nevertheless, these observations make clear how stimulating and insightful new BP studies have been, and how greatly they are contributing to linguistics.

3 Structure and content of the manual

As already outlined above, this is the first manual of BP linguistics written in English. Our aim has been to give visibility to the enormous advances in the study of BP over recent decades which are either inaccessible to readers with little or no knowledge of Portuguese or are scattered in the specialized literature of the different sub-disciplines. This explains some of the decisions with regard to the structure of this manual compared to other recently published manuals of Portuguese linguistics, such as Martins/Carrilho (2016) or Wetzels/Costa/Menuzzi (2016). First of all, both of those manuals are largely about research on contemporary Portuguese and dedicate relatively little space to diachronic issues. However, given the importance of diachronic research in contemporary BP linguistics and its reduced visibility at an international le-

vel, this had to be taken into consideration in terms of the structure of the present volume, which gives roughly equal weight to synchrony and diachrony, at least for the core domains of grammar. It is clear that a comprehensive overview is very difficult to achieve in our highly specialized field, which is also strongly interdisciplinary by nature. Since this a first attempt for BP linguistics, we decided to structure the volume around the core external factors and internal domains that constitute language in general: contact and social factors on the external side, and lexicon, phonetics/phonology, morphology, syntax and semantics on the internal side, giving the manual a rather classical (or traditional) profile. While such a broad conceptualization of chapters makes it difficult to reflect the sometimes highly technical discussions in modern linguistic analyses, it has the advantage of giving a solid description of the essentials of BP grammar on which some degree of consensus has been reached.

In the remaining part of this section, the structure and content of this manual is briefly compared to the two aforementioned manuals. Besides focusing on EP, and being written in Portuguese, Martins/Carrilho (2016) selectively included more specialized chapters, reflecting dynamic areas of current research, such as chapters on digital humanities, the inflected gerund in dialectal Portuguese, different chapters that focus on the L1 acquisition of particular grammatical elements, and the abovementioned chapter on corpus linguistics. It also features selected interdisciplinary overviews, for instance of computational linguistics, clinical linguistics, and cognitive science. The volume provides an introduction which presents the Portuguese language in synchrony and diachrony, and it is structured in two blocks: *visões gerais* ('general perspectives') and *tópicos de sintaxe, semântica e fonologia* ('topics in syntax, semantics and phonology').

The handbook by Wetzels/Costa/Menuzzi (2016) contrasts with the present manual and the one by Martins/Carrilho in that it features a total of 32 rather short chapters, many of which have a comparative stance, contrasting EP with BP. The chapters are also quite dense in terms of data and often somewhat technical in nature, thus demanding a rather high degree of linguistic training or background from the reader and an interest in the formal analysis of linguistic structures. The handbook does not provide an introduction or further internal structuring. It opens with chapters on external factors, history and a broad comparison of word order between EP and BP. Subsequently, the chapter topics flow from phonetics and phonology to morphology and then to syntax and semantics, also addressing interface questions. The final chapters discuss "processes" understood in a broad way: processes of grammatical change, of phonological and syntactic variation, and of acquisition.

This rather short and superficial comparison already makes clear that these manuals, despite being about Portuguese linguistics, are quite complementary in content, and also with respect to their expected readership. Only together can they hope to give a somewhat comprehensive overview of the linguistic research on and in the Lusophone world.

3.1 What is not to be expected from this manual

In the previous section, we explained our decision to focus this manual on the core external factors and internal domains of BP grammar. This means, however, that it was not possible to include a wide range of further topics in Brazilian linguistics from highly active and productive subfields, specifically most inter- and cross-disciplinary approaches, and also most applied sciences. Thus, Brazilian studies on translation, grammar and dictionary writing, language teaching, computational linguistics, and others, such as forensic and clinical linguistics, are not covered here. Some recent tendencies in grammar writing, however, are briefly discussed in ↗1 Brazilian Portuguese linguistics: an overview. For a book-length historiography of Brazilian grammar writing, see Cavaliere (2014). As for lexicography, the reader might usefully consult Bugueño Miranda/Borba (2019), a fairly recent manual reflecting lexicographic work in Brazil, which also contains an overview and comparison of the different types of contemporary BP dictionaries and sets out several desiderata for further work here. The question of language teaching has always been a hot and controversial topic in Brazil, which subdivides into many underlying issues. While language teaching has habitually been based on traditional normative grammar, more recently, and as a result of decades of in-depth research into the linguistic reality of Brazil, new approaches are gaining ground and prejudices towards non-normative spoken language are being addressed, in both L1 and L2 teaching. One undoubted highlight of this tendency is the monumental *Gramática pedagógica do português brasileiro* (Bagno 2011). Outside Brazil, the needs of BP L2 and EP L2 teaching poses a series of strategic and practical questions, both for institutions offering such courses and for the learners. Such issues, among others, are addressed in Reimann et al. (in prep.).

Computational linguistics has always been an interdisciplinary field with high degree of relevance for the linguistic community, in Brazil as elsewhere. Some of the activities of computational linguistics in Brazil have already been mentioned in section 2.1, above. Pardo et al. (2010) is the latest comprehensive overview of the activities in this field that we could find. The authors report surveys regarding main research topics and groups, established conferences and journals, and also discuss challenges in the field. Given its dynamics, however, the current picture might already be somewhat different.

Having a rather traditional/classical profile, the largest unit of linguistic analysis addressed in this manual, the text, is treated in terms of the linguistic consequences of its modality (spoken/written) and in terms of text type or discourse tradition (↗17 Spoken *vs.* written language). It has to be acknowledged, however, that beyond this limit there exists a vast and very broad area of research, including several subfields, dedicated to the study of text, and hence, in a broader sense, this is also considered as being part of linguistics. This area spans work on conversational analysis to different flavors of discourse analysis. For the latter, a genuinely Brazilian school has emerged over recent decades. As explained in more detail in ↗1 Brazilian Portuguese linguis-

tics: an overview, conversational analysis entered the scene largely as an auxiliary science for the large sociolinguistic projects, in which oral conversations were recorded for analysis. Yet once the value of conversational analysis for the description of spoken language data had become apparent, the first volume of the *Gramática do português culto falado no Brasil* ('Grammar of Brazilian Portuguese speakers with a higher education') was dedicated entirely to this kind of analysis (Jubran/Koch 2006). There is also a continuous exchange of ideas and findings between conversational analysis and text linguistics (Bentes/Leite 2010). Discourse analysis goes beyond structure and content and seeks to address the ideologies behind texts, looking at how such ideologies are conveyed and how they (may) affect society. This perspective on the study of text has found many followers in Brazil and has led to the establishment of a Brazilian school, including different orientations. On the one hand, some research leans towards the originally French tradition, as found, for instance, in the work of Michel Foucault; on the other hand, other work draws on the British tradition, inspired by the writing of Norman Fairclough. The evolution of the former school across three subsequent phases is described, for instance, in Paula/Stafuzza (2010). As to the latter, British-influence school, and also in its evolution into critical discourse analysis (CDA), see Ramalho/Resende (2011). Besides having applied the respective research programs which originally inspired them, these schools should also be credited for having translated the work of their "founding fathers", making directly available to a Brazilian readership the ideas of Foucault, Fairclough, van Dijk and others. This also holds true for the work of Mikhail Bakhtin and Valentin Voloshinov and their overall presence in Brazilian studies in the field of discourse and dialogue analysis, as well as in research on "discourse genres" (on the reception of Bakhtin in Brazil in general, see Brait 2005).

Another huge topic of importance that did not fit the profile of our volume is the Brazilian sign language LIBRAS (*língua brasileira de sinais*), another Brazilian natural language in its own right, whose study, however, has often been undertaken by people from other fields, such as cognitive neuroscience, rather than linguistics proper. This, for instance, has been the case with important lexicographic resources,[13] including the most comprehensive modern dictionary (Capovilla et al. 2017), which comprises 14,500 sings and also contains detailed information on morphology, etymology and regional use. The first graduate program in Brazilian sign language within a philological department (*Letras-Libras*) was inaugurated at the Federal University of Santa Catarina in 2006 (Quadros/Stumpf 2009), an initiative subsequently replicated in many other states. For an introduction to Libras linguistics, consider Rodrigues/Va-

13 The journal of São Paulo University published an interesting interview regarding the development of these resources: https://jornal.usp.br/cultura/dicionario-da-lingua-de-sinais-exigiu-25-anos-de-pesquisas/, last accessed 11.02.2022.

lente (2011); current developments are presented and discussed in Quadros (2020), a collective volume published in English.

3.2 Summary of the following chapters

The contributions to this handbook were written by specialists in their respective fields, with a strong presence of Brazilian scholars, and with aim of making the aforementioned rich production of BP linguistics accessible to an international public. To re-appropriate Labov's principle (formulated previously by the Neogrammarians), we believe that we should 'use the past to explain the present', and hence this handbook has a historical as well as a contemporary-synchronic focus, which also constitutes its main internal structure. Within each block, we start with the presentation of social context and external factors, then move from smaller units of linguistic structure to the larger ones—from the sound inventory to words, phrases and sentences. Finally, the lexicon as a proper structural domain is presented, and in the synchronic block some further topics are also covered. Several of the papers are co-authored by people with different theoretical background, thus giving insights into the collaborative spirit of Brazilian linguistics noted above.

The overview by Cristina Altman and Ataliba Teixeira de Castilho (↗1 Brazilian Portuguese linguistics: an overview) sheds lights on important moments across five centuries of linguistic thought and investigation in Brazil, tracing its evolution from the very beginning in the 16th century to recent large, interdisciplinary projects and modern linguistics, also including a personal outlook. It is followed by a diachronic block of six chapters. In chapter 2, (↗2 The social history of Brazilian Portuguese), Tânia Lobo describes the drastic changes that have occurred in Brazilian society, from the first colonial settlements through centuries of colonialism and slavery, to the immense growth of population and immigration in the 20th century, all with major consequences for the evolution of the language(s) at different points in time. In chapter 3, Dante Lucchesi and Alan Baxter then account for these centuries of intense linguistic contact of Portuguese with many American and African languages (↗3 The history of linguistic contact underlying Brazilian Portuguese) and provide insights for contact linguistics from the complex Brazilian scenario. The chapters describing the internal evolution of BP begin with phonetics and phonology (↗4 Historical phonetics and phonology), a contribution by Volker Noll. The author also makes use of illustrative and systematic comparisons of the phonetic and phonological system of EP. In ↗5 Historical morphology, Mário Viaro addresses the challenges of historical morphology in BP, describing significant evolutions in word formation but also conceptual issues that have been addressed in this domain of investigation. In ↗6 Historical syntax, the current state of investigation with regard to historical syntax is addressed, combining the generative/formal and functional perspectives, in a collaborative study by Charlotte Galves and Célia Lopes. The focus is on evolutions from the 19th century to present. Américo Venâncio

Machado Filho's study ↗7 The history of the lexicon, first describes the most important current research activities in this field of research and some theoretical and terminological issues, before giving a comprehensive overview of the evolution of the BP lexicon with a focus on the many borrowings from different languages through the centuries. ↗8 The debate on Brazilian and European Portuguese, which closes the "historical" block, is not a diachronic account in the narrow sense, but presents the history of the debate on the relationship between EP and BP and the different kinds of arguments and proposals that have been endorsed at different points in time. This also includes arguments based on the modern conception of the notion of grammar.

The synchronic block starts with a description of variation in contemporary BP—first from a dialectological perspective, in a study by Vanderci Aguilera in ↗9 Dialectology and linguistic geography. She also gives an account of the evolution of this sub-area and presents many of the main geolinguistic resources that have been built, including the first two published volumes of the first comprehensive linguistic atlas of Brazil. After this, Uli Reich and Ronald Beline provide an overview of the different phases of one of the most dynamic and active research areas whose contribution to our current understanding of BP cannot be overestimated—its sociolinguistics, in ↗10 Sociolinguistics. John Lipski completes the block, taking more external perspectives in ↗11 Brazilian Portuguese: contemporary language contacts, a detailed description of contact scenarios and their consequences along the immense border of Brazil, mostly with Spanish-speaking countries, and also of linguistic contact due to waves of migration to Brazil in modern times. ↗12 Sound-related aspects of Brazilian Portuguese by Dermeval da Hora and Elisa Battisti describes the current state of BP phonetics and phonology, including also aspects of prosody. Chapter 13, by Carlos Alexandre Gonçalves, is dedicated to lexical and inflectional morphology in the synchrony of BP and focuses on some important typological particularities, among others in the field of word formation (↗13 Morphology). Chapter 14, on syntax, is another collaborative chapter, a joint study by Sanderléia Longhin and Maria Clara Paixão de Souza, and offers both a functional and a formal view on several issues in current BP syntax (↗14 Syntax). The BP lexicon is presented in chapter 15 by Maria da Graça Krieger, with insights into BP lexicology, lexicography and studies on terminology (↗15 Lexicon). Chapter 16 is dedicated to semantics. Rodolfo Ilari and Renato Basso (↗16 On the history of semantics studies in Brazil) provide an overview of what is presented as "semantics" in the panorama of BP linguistics, and include observations on "implicit" studies of semantic issues, as well as offering some perspectives for possible evolutions in BP semantic research.

As mentioned above, the tension between the spoken and the written language is an important issue throughout the history of BP, and its differences are relevant for many chapters of the book. However, Alessandra Castilho da Costa's chapter (↗17 Spoken *vs.* written language) is dedicated exclusively to this issue, not from the viewpoint of a particular linguistic phenomenon, but rather considering language from the point of view of the communicative situations determining its production and from

the medium of its realization. Patricia Carvalinhos' contribution (↗18 Onomastics and toponomastics) considers the history and the present state of proper names (toponyms and anthroponyms) and shows how names reflect cultural history (including through the omission of important parts of this history) as well as the current multifaceted dynamism of naming. José Luiz Fiorin (↗19 Linguistic policy and the orthographic agreement) gives an overview of implicit and explicit language policies concerning the national language and the languages within Brazil, as well as the relation between BP and other varieties in the Lusophone world, namely in the field of orthography. The final chapter, by Elisângela Teixeira and Michele Calil Alves, presents the state of the art and some selected studies in the rather recent discipline of psycholinguistics in Brazil, with a focus on processing and acquisition of BP structures (↗20 Psycholinguistic studies: language acquisition and processing).

4 Final remarks

In the tradition of Romance linguistics, colonial varieties used to be considered as somewhat uniform as compared to the rich diatopic diversity of "primary dialects" (Coseriu 1980) that can be found in the European motherlands. This is also an obviously Eurocentric perspective that can be explained by the detailed knowledge of language variation in Europe due to fine-grained dialectological studies since the 19th century and by the lack of systematic knowledge of language situations overseas. If we look at Brazil from inside, a country of almost continental dimensions and with a complex linguistic history, we discover a fascinating reality which incorporates a whole cosmos of phenomena, and we see that there exists a flourishing activity of linguistic studies in all imaginable fields.

We hope that this manual will contribute to a better knowledge of this reality, and will make possible a vision of Brazilian Portuguese without any colonial bias. This is not an ideological perspective, and by no means does it seek to deny the relationship between current BP and its colonial and post-colonial past. Indeed, a significant part of this manual is dedicated to a historical view. Yet this historical background should above all serve as a basis for a differentiated consideration of the many contemporary insights from diverse perspectives as presented in this book.

5 References

Avelar, Juanito Ornelas de (2009), Ter, ser *e* estar. *Dinâmicas morfossintáticas no português brasileiro*, Campinas, Editora RG.

Avelar, Juanito Ornelas de (2018), *Sentenças possessivas e existenciais*, in: Sonia Cyrino/Maria A. Torres Morais (edd.), *História do português brasileiro*, vol. 6: *Mudança sintática do português brasileiro: perspectiva gerativista*, São Paulo, Contexto, 72–149.

Bagno, Marcos (2011), *Gramática pedagógica do português brasileiro*, São Paulo, Parábola.

Barbosa, Afranio (2019), *A plataforma de corpora do PHPB: uma apresentação ad infinitum*, in: Ataliba T. de Castilho (ed.), *História do português brasileiro*, vol. 2: *Corpus diacrônico do português brasileiro*, São Paulo, Contexto, 16–67.

Bentes, Anna Christina/Leite, Marli Quadros (edd.) (2010), *Linguística de texto e análise da conversação: panorama das pesquisas no Brasil*, São Paulo, Cortez.

Berber Sardinha, Tony (2010), *Manual Corpus Brasileiro*, https://www.linguateca.pt/Repositorio/manual_cb.pdf, last accessed 11.02.2022.

Berber Sardinha, Tony/São Bento Ferreira, Telma de Lourdes (edd.) (2014), *Working with portuguese corpora*, London, Bloomsbury.

Brait, Beth (2005), *Bakhtin. Dialogismo e construção do sentido*, Campinas, Unicamp.

Bugueño Miranda, Félix Valentín/Borba, Laura Campos de (2019), *Manual de (Meta)Lexicografia,* Goiânia, Espaço Acadêmico.

Calindro, Ana Regina (2020), *Ditransitive constructions: What sets Brazilian Portuguese apart from other Romance languages?*, in: Anna Pineda/Jaume Mateu (edd.), *Dative constructions in Romance and beyond*, Berlin, Language Science Press, 75–95.

Capovilla, Fernando César, et al. (2017), *Dicionário da Língua de Sinais do Brasil: A Libras em Suas Mãos*, 3 vol., São Paulo, USP.

Castilho, Ataliba T. de (ed.) (2019), *História do português brasileiro*, vol. 2: *Corpus diacrônico do português brasileiro*, São Paulo, Contexto.

Castilho, Ataliba T. de, et al. (2019), *Diacronia da concordância*, in: Ataliba T. de Castilho (ed.), *História do português brasileiro*, vol. 5: *Mudança sintática das construções: perspectiva funcionalista*, São Paulo, Contexto, 284–399.

Cavaliere, Ricardo (2014), *A gramática no Brasil. Ideias, percursos e parâmetros*, Rio de Janeiro, Lexikon.

Cépeda, Paola/Cyrino, Sonia (2020), *Putting objects in order: Asymmetrical relations in Spanish and Portuguese ditransitives*, in: Anna Pineda/Jaume Mateu (edd.), *Dative constructions in Romance and beyond*, Berlin, Language Science Press, 97–116.

Collischonn, Gisela/Wetzels, Leo W. (2016), *Syllable Structure*, in: Leo W. Wetzels/João Costa/Sérgio Menuzzi (edd.), *The Handbook of Portuguese Linguistics*, Malden/Oxford, Wiley, 86–106.

Coseriu, Eugenio (1980), *"Historische Sprache" und "Dialekt"*, in: Joachim Göschel/Pavle Ivic/Kurt Kehr (edd.), *Dialekt und Dialektologie. Ergebnisse des internationalen Symposiums "Zur Theorie des Dialekts",* Wiesbaden, Steiner, 106–115.

Costa, João (2011), *Topic prominence is not a factor of variation between Brazilian and European Portuguese*, in: Janine Berns/Haike Jacobs/Tobias Scheer (edd.), *Romance Languages and Linguistic Theory 2009: Selected papers from "Going Romance" Nice 2009*, Amsterdam/Philadelphia, Benjamins, 71–88.

Cresti, Manuela/Moneglia, Massimo (ed.) (2005), *C-ORAL-ROM. Integrated reference corpora for spoken Romance languages*. Amsterdam/Philadelphia, Benjamins.

Cruz, Marisa/Frota, Sónia (2014), *Rhythm in central-southern varieties of European Portuguese: production and perception*, in: António Moreno et al. (edd.), *Textos Selecionados. XXIX Encontro Nacional da Associação Portuguesa de Linguística*, Porto, Associação Portuguesa de Linguística, 214–230.

Cyrino, Sonia/Nunes, Jairo/Pagotto, Emilio (2009), *Complementação*, in: Mary A. Kato/Milton do Nascimento (edd.), *Gramática do português culto falado no Brasil*, vol. 3: *A construção da sentença*, Campinas, Unicamp, 47–96.

Davies, Mark (2016–), *Corpus do Português: Web/Dialects*, http://www.corpusdoportugues.org/web-dial/, last accessed 11.02.2022.

Davies, Mark/Ferreira, Michael (2006–), *Corpus do Português: Historical Genres*, http://www.corpusdoportugues.org/hist-gen/, last accessed 11.02.2022.

Duarte, Maria Eugênia Lammoglia (2012), *O sujeito em peças de teatro (1833–1992): estudos diacrônicos*, São Paulo, Parábola.

Estivalet, Gustavo Lopez (2015), *Manual do Léxico do Português Brasileiro–LexPorBR: corpus psicolinguístico*, http://www.lexicodoportugues.com/downloads/manual_lexporbr_alfa2_pt-br.pdf, last accessed 11.02.2022.

Frota, Sónia/Vigário, Marina (2001), *On the correlates of rhythmic distinctions: The European/Brazilian Portuguese case,* Probus 13, 247–275.

Galves, Charlotte/Andrade, Aroldo Leal de/Faria, Pablo (2017), *Tycho Brahe Parsed Corpus of Historical Portuguese,* http://www.tycho.iel.unicamp.br/~tycho/corpus/texts/psd.zip, last accessed 11.02.2022.

Jubran, Clelia Spinardi/Koch, Ingedore Grunfeld Villaça (edd.) (2006), *Gramática do Português Culto Falado no Brasil, vol. 1: Construção do texto falado*, Campinas, Universidade Estadual de Campinas/Fundação de Amparo à Pesquisa do Estado de São Paulo.

Kabatek, Johannes (2007), *Existe um ciclo de gramaticalização do artigo na România?*, in: Jânia Ramos (ed.), *Para a história do português brasileiro*, vol. 5: *Estudos sobre mudança lingüística e história social*, Belo Horizonte, FALE/UFMG, 13–51.

Kewitz, Verena/Simões, José da Silva (2019), *Características e potencialidades dos corpora do português paulista*, in: Ataliba T. de Castilho (ed.), *História do português brasileiro*, vol. 2: *Corpus diacrônico do português brasileiro*, São Paulo, Contexto, 244–289.

Kewitz, Verena, et al. (2018), *As preposições: Aspectos históricos e usos atuais*, in: Célia R. Lopes (ed.), *História do português brasileiro*, vol. 4: *Mudança sintática das classes de palavra: perspectiva funcionalista*, São Paulo, Contexto, 294–385.

Lemle, Miriam/Naro, Anthony J. (1977), *Competências básicas do português*, Rio de Janeiro, Mobral/Fundação Ford.

Li, Charles N./Thompson, Sandra A. (1976), *Subject and Topic: A New Typology of Language*, in: Charles N. Li (ed.), *Subject and Topic*, New York, Academic Press, 457–489.

Lopes, Célia Regina (ed.) (2018), *História do português brasileiro*, vol. 4: *Mudança sintática das classes de palavra: perspectiva funcionalista*, São Paulo, Contexto, 294–385.

Martins, Ana Maria (2016), *Introdução: O português numa perspetiva diacrónica e comparativa*, in: Ana Maria Martins/Ernestina Carrilho (edd.), *Manual de linguística portuguesa*, Berlin/Boston, De Gruyter, 1–39.

Martins, Ana Maria/Carrilho, Ernestina (edd.) (2016), *Manual de linguística portuguesa*, Berlin/Boston, De Gruyter.

Martins, Marco Antonio (2018), *A sintaxe dos pronomes pessoais clíticos na história do português brasileiro*, in: Sonia Cyrino/Maria A. Torres Morais (edd.), *História do português brasileiro 6: Mudança sintática do português brasileiro: perspectiva gerativista*, São Paulo, Contexto, 150–209.

Massini-Cagliari, Gladis/Cagliari, Luiz Carlos/Redenbarger, Wayne J. (2016), *A Comparative Study of the Sounds of European and Brazilian Portuguese. Phonemes and Allophones*, in: Leo W. Wetzels/João Costa/Sérgio Menuzzi (edd.), *The Handbook of Portuguese Linguistics*, Malden/Oxford, Wiley, 56–68.

Mendes, Amália (2016), *Linguística de corpus e outros usos dos corpora em linguística*, in: Ana Maria Martins/Ernestina Carrilho (edd.), *Manual de linguística portuguesa*, Berlin/Boston, De Gruyter, 224–251.

Modesto, Marcello (2001), *As construções clivadas no português do Brasil: relações entre interpretação focal, movimento sintático e prosódia*, São Paulo, Humanitas.

Noll, Volker (2008), *O português brasileiro. Formação e contrastes*, São Paulo, Globo.

Nunes, Maria das Graças Volpe, et al. (1996), *A construção de um léxico para o português do Brasil: lições aprendidas e perspectivas*, in: *Anais do II Encontro para o processamento de português escrito e falado*, Curitiba, CEFET-PR, 61–70.

Paixão de Souza, Maria Clara (2010), *A morfologia de flexão no português do Brasil: Ensaio sobre um discurso de perda*, Estudos da Lingua(gem) 8, 55–82.

Pardo, Thiago A. S, et al. (2010), *Computational linguistics in Brazil: an overview*, in: *Proceedings of the NAACL HLT 2010 Young Investigators Workshop on Computational Approaches to Languages of the Americas*, Los Angeles, Association for Computational Linguistics, 1–7.

Paula, Luciane de/Stafuzza, Grenissa (edd.) (2010), *Da análise de discurso no Brasil à análise de discurso do Brasil: três épocas histórico-analíticas*, Uberlândia, UFU.

Pontes, Eunice (1987), *O tópico no Português do Brasil,* Campinas, Pontes.

Quadros, Ronice Müller de (ed.) (2020), *Brazilian Sign Language Studies*, Berlin/Boston, De Gruyter Mouton.

Quadros, Ronice Müller de/Stumpf, Marianne Rossi (2009), *O primeiro curso de graduação em letras língua brasileira de sinais: educação a distância*, ETD – Educação Temática Digital 10/2, 169–185.

Ramalho, Viviane/Resende, Viviane de Melo (2011), *Análise de discurso (para a) crítica: o texto como material de pesquisa*, Campinas, Pontes.

Raso, Tommaso/Mello, Heliana (2012), *C-ORAL-BRASIL I: Corpus de referência do português brasileiro falado informal,* Belo Horizonte, Editora UFMG.

Reimann, Daniel, et al. (in preparation), *Fachdidaktik Portugiesisch*, Tübingen, Narr.

Rodrigues, Cristiane Seimetz/Valente, Flávia (2011), *Aspectos Linguísticos da Libras*, Curitiba, IESDE Brasil.

Rosemeyer, Malte (2018), *PorThea. A historical corpus of Portuguese theater plays*, http://www.romanistik.uni-freiburg.de/rosemeyer/05corpus.html, last accessed 11.02.2022.

Silva, Andressa Christine Oliveira da/Fonseca, Aline Alves (2018), *Prosody and Processing. Comprehension and Production of Topic-Comment and Subject-Predicate Structures in Brazilian Portuguese*, Revista de Estudos da Linguagem 26/4, 1601–1646.

Silva, Edila Vianna da (2015), *Bancos de dados sociolinguísticos em português*, Idioma 29, 168–180.

Simões, José da Silva (2007), *Sintaticização, discursivização e semanticização das orações de gerúndio no português brasileiro*, Doctoral thesis, São Paulo, Universidade de São Paulo.

Simões, José da Silva (2012), *A importância da história dos gêneros para a constituição de corpora diacrônicos*, in: Marco Antonio Martins/Maria Alice Tavares (edd.), *História do Português Brasileiro no Rio Grande do Norte. Análise linguística e textual da correspondência de Luís da Câmara Cascudo a Mario de Andrade, 1924–1944*, Natal/RN, Universidade Federal do Rio Grande do Norte, 55–94.

Teixeira de Souza, Lilian (2015), *Three types of negation in Brazilian Portuguese*, Lingua 159, 27–46.

Vasco, Sérgio Leitão (2006), *Construções de tópico na fala popular*, Doctoral thesis, Rio de Janeiro, Universidade Federal do Rio de Janeiro.

Wall, Albert (2017), *Bare nominals in Brazilian Portuguese. An integral approach*, Amsterdam/Philadelphia, Benjamins.

Wetzels, Leo W./Costa, João/Menuzzi, Sérgio (edd.) (2016), *The Handbook of Portuguese Linguistics*, Malden/Oxford, Wiley.

Zampieri, Marcos/Becker, Martin (2013), *Colonia: Corpus of Historical Portuguese*, in: *ZSM Studien*, vol. 5: *Special Volume on Non-Standard Data Sources in Corpus-Based Research*, Aachen, Shaker.

Cristina Altman and Ataliba T. de Castilho

1 Brazilian Portuguese linguistics: an overview

Abstract: The present overview encompasses five centuries of Brazilian linguistic historiography. It opens with the descriptive strategies used by the Jesuits in the description of the native languages in colonial times and outlines the emergence of two practices of linguistic description in Brazil, sponsored by two research centres in 19th century Brazil, the Instituto Histórico Geográfico Brasileiro and the Colégio Dom Pedro II. It also summarizes the controversy surrounding Brazilian and European Portuguese, plus the institutionalization of a linguistic discipline as autonomous from traditional philology and dialectology, together with some notable results of this. These include the description of spoken Brazilian Portuguese, new tendencies in grammar writing, and the renewed interest in the history of Brazilian Portuguese. Finally, an outlook on new issues in Brazilian linguistics is offered.

Keywords: linguistic historiography, grammaticography, Amerindian languages, missionaries, institutionalization, history of Brazilian Portuguese, grammar of spoken Brazilian Portuguese

1 'The most commonly used language along the Brazilian coast'

Although Brazil is quite heterogeneous in its constitution as a nation, Brazilian linguistic thought developed around the idea of the uniqueness of the country's languages, Tupinambá in colonial times, and Brazilian Portuguese (BP) after its independence. As a territory coveted by the Portuguese, the Spanish, the French, and the Dutch, Brazil was, during the centuries in which it was a colony of Portugal (1500–1822), one of the key elements in the military, commercial and missionary Iberian enterprise. As a consequence, we face the initial paradox of describing "Brazilian" linguistic thought far earlier than we can talk of a Brazilian tradition in the study of languages, and well before we can consider the peculiarities of the Brazilian languages themselves.

The belief that we speak one single language, from the North to the South, has become a cliché since Joseph of Anchieta (1534–1597) wrote his grammar on 'the most commonly used language along the Brazilian coast'[1] (Anchieta [7]1990, [1]1595). The names associated with this language—Tupinambá, or the ancient Tupi—and the gen-

1 "a lingoa mais usada na costa do Brasil".

https://doi.org/10.1515/9783110405958-002

eral languages derived from it (cf. Rodrigues 1996; 2000), as well as the functions and values aggregated to them during the subsequent centuries, differ through the literature (Rodrigues 1996; Rosa 2003; Altman 2007). Yet the strategy of choosing a single variety, to the detriment of hundreds of others, as the supra-regional means of communication and codification, lasted at least until the end of the 19th century, if not later. It was maintained even after the native peoples "ascended" to the category of free citizens, and the Portuguese language was decreed the official language of the country (cf. the Directory of 1787, of the Marquis of Pombal).

From our current vantage point, we can say that ancient Tupi and Tupinambá, spoken in the 16th and the 17th centuries, the Paulista General Language, in the 17th and 18th, and the Amazonian General Language, Nheengatú, spoken from the 17th century onwards (Bessa Freire 2004), are all geographical and historical varieties of the same subset of the Tupi-Guarani family, the one disseminated along the Atlantic coast, from the South upwards as far as the Amazon region. However, the colonial sociolinguistic reality was, of course, much more complex than that. There were hundreds of other languages and ethnicities, not to mention the various African languages that entered Brazil with the slave trade, in successive waves, from the middle of the 16th century to the end of the 19th (Bonvini 2008).

2 The missionary enterprise and the general languages

The arrival of the first missionaries made possible the establishment of European control over the territory—thus guaranteeing to their sponsors the economic and commercial exploitation of the land—leading to conflicts between natives and colonists and helping to define, in practice, the nation to which the "discovered" territory belonged: Spain or Portugal. At the same time, missionaries also became responsible for everything related to education, language description and language teaching in Brazil for almost three centuries.

The Franciscans were the first to arrive, as early as 1515 (Jancsó et al. 1994, 31), but with the exception of André Thévet (1502–1590), who published his *Cosmographie Universelle* in 1575, their writings were not as numerous and their influence was not as significant as those of the Jesuits. French Calvinists also tried to establish their own territory in Rio de Janeiro, founding what they called the Antartic France (a short-lived experience lasting only from 1555 until 1560). The Calvinist priest Jean de Léry (1534–1611) was the first to register forms of the Tupi language, quite probably as it was spoken at that time, as noted by Rodrigues (1997), in that it was not a translation of a European text but the reproduction of a dialogue between a native and a Frenchman (Léry 1980, 275–303). Lery's work, however, is not a grammar and from what is left of it, following several (not always careful) transcriptions through the centuries, we can

hardly say that we have before us an attempt to describe the Tupi language—or at least some part of it—but rather a casual collection of colloquial phrases.

No other religious order exerted as much power and provided as much information about Portuguese America as that of the Jesuits, who first arrived in Brazil in 1549. They dominated the religious and educational agenda in Brazil until 1759, when they were expelled from Portugal and its colonies, and Portuguese was declared the only official language of Brazil. The Jesuits were responsible for the first grammatical descriptions and dictionaries of the ancient languages spoken in Brazil. The Tupi language was described twice, by the Spanish-born Jesuit Joseph de Anchieta (1534–1597) and by the Portuguese Jesuit Luís Figueira (1575–1643), (cf. Anchieta [7]1990, [1]1595; Figueira 1621). *Kiriri*, a language which became extinct by the 18th century, was described by the Jesuit Luís Mamiani (1652–1730) (cf. Mamiani 1699), and we must also mention the Jesuit Pedro Dias' (1622–1700) grammar of Kimbundu, although it concerns a non-native African language spoken in Brazil in the 17th century (cf. Dias 1697).

The main motivation of the European missionaries for bringing these languages into some form of grammatical description was the preparation and instruction of the missionaries themselves, rather than for the natives. This is why they were written in Portuguese (and Latin, in the case of Anchieta) and, like many others contemporary grammars, they followed the characteristic Latin-oriented pattern of the time: rules for the pronunciation of the language; rules of declension for nouns, adjectives and pronouns; rules for the formation and conjugation of verbs; and a variable list of the other 'parts of speech': prepositions, adverbs, interjections and conjunctions. The 'prepositions', were, in fact, postpositions, as Figueira himself recognizes:

> 'All the prepositions of this language might better be called postpositions, because they are always placed after the noun which they govern' (Figueira 1621, 65r).[2]

The most important thing to observe, however, is that these grammars not only register the languages they described, but are also significant testimonies of the Portuguese language (and of the metalanguage of description) used in the 16th and 17th centuries.

Observe the example:

> 'As well as in the Portuguese language, instead of cases we add some prepositions to the nouns, v.g. Pedro, of Pedro, to Pedro, for Pedro, with Pedro, &c., so in this language any substantive noun is governed, & varies with prepositions' (Figueira 1621, 3v).[3]

---

2 "Todas as prepoſições deſta lingua, ſe podem melhor chamar poſpoſições, por que ſempre ſe poem deſpois do nome, que regem."

3 "Aſſí como na lingua Portugueſa em lugar de caſos ajuntamos algũas prepoſições aos nomes, v.g. Pedro, De Pedro, A Pedro, Pera Pedro, Com Pedro, &c. Aſſi tambem neſta lingua qualquer nome ſuſtantiuo he gouernado, & varia com prepoſições."

This excerpt illustrates the intermediation of the Portuguese language in the description of Tupinambá: as well as having discovered that nouns in Romance languages substituted the inflected Latin forms, provoking a re-evaluation of the constructions with prepositions (Robins 1967, 79–80), nouns in Tupinambá, through formal analogy, led to a re-evaluation of postpositions.

In addition to the catechisms, grammars, glossaries and vocabularies provided by the Jesuits, colonial linguistic documentation includes an enormous variety of religious and non-religious sources, which we can generally refer to as travel reports. Some of these took the form of longer descriptive essays in which sparse linguistic data and personal opinions about the native languages are mixed with ethnological and geographical information, the texts by Pero de Magalhães Gândavo (1858, [1]1576), and Gabriel Soares de Souza's impressive *Tratado descriptivo do Brasil em 1587* (Souza [3]1938, [1]1587) being good examples. Both were colonists who, although non-religious, must have had some formal education. The example below gives a taste of the kind of linguistic description present in these materials:

> 'They [the natives] are very gracious when they speak, especially the women; they are very compendious in the form of language, and very copious in their prayer; but they lack three letters of the A B C, which are F, L, R capital or double, quite a remarkable thing; because if they do not have F, it is because they do not have faith in anything they adore; [...]. And if they do not have L in their pronunciation, it is because they do not have any law to observe, neither guideline to govern themselves; [...]. And if they do not have R in their pronunciation, it is because they have no *rex* [king] to rule them, and whom to obey [...]' (Souza [3]1938, [1]1587, 364–365).[4]

The more obvious point of interest in this type of source lies in the list of terms and translations that they eventually present; the less obvious, but not less important, is the sort of parallel linguistic information they carry, in that it allows us to better perceive the beliefs, values, and attitudes of the European colonizer toward the native peoples—and their languages.

3 The *IHGB* and the *Colégio Pedro II*: the institutionalization of two practices

The 17th and 18th centuries in Brazil witness the rise to significance of the Amazonian General Language in the colonization of the North of the country, but, from the point

4 "Tem muita graça quando fallam, mórmente as mulheres; são mui compendiosas na fórma da linguagem, e muito copiosas no seu orar; mas falta-lhes tres letras do A B C, que são F, L, R grande ou dobrado, cousa muito para se notar; porque senão tem F, é porque não tem fé em nenhuma cousa que adorem; [...]. E se não tem L na sua pronunciação, é porque não tem lei alguma que guardar, nem preceitos para se governarem; [...]. E se não tem esta letra R na sua pronunciação, é porque não tem rei que os reja, e a quem obedeçam, [...]".

of view of the construction of Brazilian linguistic thought, it is the 19th century that provides the favorable conditions for the institutionalization of Brazilian identity and everything that it involves. In fact, the Second Empire (1840–1889) sees the delineation of the geographical and political borders of the country; the revival of interest in its colonial history; the resolution once and for all of the so-called "indigenous issue", i.e., that of the definitive integration of indigenous people into civil society and the establishment of a representative literature in its official language, Brazilian Portuguese.

Activities leading to the implementation of policies for the native peoples and their languages here are in great part organized by the *Instituto Histórico e Geográfico Brasileiro* (IHGB), created in 1838 by the Brazilian Emperor, D. Pedro II (1825–1891), with the task of supporting systematic research into Brazilian history, the country's inhabitants and its territory, and to promote regular meetings about its associates' findings. The creation of IHGB sparks a burst of interest in Brazilian languages, leading to the re-editing of classic texts, grammars and dictionaries, besides the work of various colonial chroniclers, which were also re-edited and published in the journal of the Institute. The interest in rescuing the Amerindian literature from previous centuries was twofold: it provided a noble and epic past for the young independent nation, and also made it possible to retrieve the indigenous vocabulary that had contributed to the formation of the BP lexicon. Assuming that native words had been altered over time, the revival of colonial literature had the advantage of recovering these words in their primitive form.

In addition, the IHGB sponsored a wave of scientific expeditions that produced original materials, such as those involving Karl Friedrich Philipp von Martius (1794–1868), Batista Caetano Nogueira de Almeida (1826–1882), José Vieira Couto de Magalhães (1837–1898), Charles Hartt (1840–1878), Karl von den Steinen (1855–1929) and Paul Ehrenreich (1855–1914). Although most of these materials deal with the Amazonian General Language,[5] they often provide data about spoken BP as well. In this sense, Couto de Magalhães' 1876 grammar of Nheengatú is illustrative:

> 'It is this Tupian construction that modified the Portuguese spoken by the people from the North of the Empire, mostly [spoken] in the one of the Amazon provinces, where they say very generally: *é melhor de você*, instead of saying: *é melhor do que você*' (Couto de Magalhães 1876, 79).[6]

Not by chance, then, the IHGB would soon become a pivotal institution, not only for the implementation of the strategies of the government regarding the native peoples, but also for the organization of those domains perceived as related: geography, his-

5 The only exception, perhaps, is Martius' (1969, [1]1867) glossary, which is, according to Rodrigues (1994; 1996), the only document to describe the southern Tupi.

6 "É esta construcção tupi que alterou o portuguez fallado pelo povo do norte do imperio, sobre tudo pelo da provincia do Amazonas, o qual diz muito commummente: *é melhor de você*, em vez de dizer: *é melhor do que você*."

tory, archaeology and ethnography. A Brazilian field linguistics emerged at this time, not as an autonomous discipline, but as a research area related to historical studies on the native peoples. This form of linguistics was to expand in the following years, quite apart from the more prestigious field of Portuguese philology, taking root in the *Colégio Pedro II*, a high school founded in Rio de Janeiro in 1837–1838.

The *Colégio Pedro II* centralized the educational system of 19th century Brazil. Its faculty was responsible for prescribing the educational programs to be followed by the few official schools that existed in the country at the time and, more importantly, for determining what should or should not be studied in linguistic matters. We should also note that higher-level studies in languages and literatures would only be developed in the 20th century. In this way, the *Colégio Pedro II* was the institution that organized teaching, fostered some research, and was responsible for publications in Portuguese for almost a century.

Divided into two, the recently inaugurated field of language studies in Brazil was hereby also to split into two kinds of practices: one seen as more practical, not requiring specific education, and related to field work on the native languages; the other, seen as more theoretical—and "more scientific"—related to the work of philologists on the Portuguese language.

4 The Brazilian (Portuguese) language

4.1 The Lexicon

From Pedra Branca's 1824–1825 (*alias*, Domingos Borges de Barros) remarks in Adrien Balbi's *Atlas Etnográfico do Globo* on the peculiarities of the 'Brazilian idiom' (cf. Pinto 1978, XV), until the middle of the 20th century, the status of the Brazilian variety of Portuguese in relation to its European matrix constituted the main focus of attention for various writers and scholars who gravitated around the *Colégio Pedro II* and, later, around the *Academia Brasileira de Letras*, created in 1897. The debates at the time show the clear perception of differences between the Brazilian and the European Portuguese, especially relating to its prosody and lexicon, as well as of internal varieties of BP, due to its geographical and social particularities, including here the gap between its (popular) spoken and its (cultivated) written modes.

Throughout the 19th century, the words that were initially presented simply to illustrate the differences in meaning between European and Brazilian uses soon acquired emotional and nationalistic nuances. Some intellectuals, even when recognizing the diversification of the Portuguese spoken in Brazil, advocated the unity of the language—like Francisco Adolfo de Varnhagen (1816–1878), for example, an important figure in 19th century Brazil, monarchist, diplomat, historian—whereas others argued for the freedom of BP through breaking with its European past. The critic, lexicographer and jurist Antonio Joaquim de Macedo Soares (1838–1905), and the

celebrated Brazilian writer Jose de Alencar (1829–1877), for example, were notorious supporters of the second position.

In his 1880 text *Algumas palavras africanas introduzidas no português que se fala no Brasil* 'Some African words introduced into the Portuguese spoken in Brazil', Macedo Soares remarked that:

> 'We have many times, in the course of writing this work, spoken of the *Brazilian dialect*. It should be noted that we did not investigate the scientific value of the word *dialect*. If we understand by dialect the word derived from the language spoken in a nation, and characteristic to a city or a province[,] *Brazilian* is not a dialect of Portuguese. On the other hand, the expression *Brazilian language* seems to us too pretentious, if one wants to use it to distinguish the spoken Portuguese in Brazil [...]. We employ as a more modest option the other expression [i.e. Brazilian dialect], which is at the same time less incorrect, and indirectly suggests that we understand that we refer to the dialectal movement that is visibly in operation in the national language' (Macedo Soares 1880, *apud* Pinto 1978, 46).[7]

This impressionistic way to deal with language diversity was reflected in the countless names given to BP. Coelho (2012, 201), by emphasizing the conceptual uncertainty at the time regarding the status of the Brazilian variety, lists all the terms used by Macedo Soares to refer to it: *Brazilian dialect*, *Brazilian Portuguese*, *national language*, *Brazilian language*, *Portuguese spoken in Brazil*, *Portuguese that is nowadays spoken in Brazil*, *Luso-Brazilian*, *our dialect*, *mother tongue*, *Luso-Brazilian dialect*, *our actual language*, *Brazilian*, *American Portuguese*, *language from this land*, *our language*, *national dialect*.

Whatever the name, the "separationists", although claiming autonomy for the Brazilian variety, recognized the authority of Portuguese authors and accepted their grammatical rules. Note in this respect Alencar's reference to P. Antonio Vieira (1608–1697), when defending himself of the misuse of the verb *reclinar* 'to recline':

> 'Will there be an impediment for me to do with the verb *reclinar* [recline] what Vieira did with the verb *inclinar* [incline], so related and close to each other? Does the diversity of the prefix modify the word so much that it changes its nature? Or will it be that we Brazilians only have the right to coin words from *Tupi*, like *cuia* and *tiquara*, it being forbidden for us to touch the holy ark of classicism?' (Alencar 1873 *apud* Pinto 1978, 110).[8]

---

7 "Temos muitas vezes, no correr deste escrito, falado em *dialeto brasileiro*. Cumpre observar que não apuramos o valor científico da palavra *dialeto*. Se entendermos por dialeto a palavra derivada da língua geral de uma nação e particular a uma cidade ou província o brasileiro não é dialeto do português. Por outro lado, a expressão *língua brasileira* nos parece demasiado pretensiosa, se se quer com ela distinguir o português falado no Brasil [...]. Empregamos como mais modesta a outra expressão [i.e. dialeto brasileiro], que é ao mesmo tempo menos incorreta, e dá bem a entender que nos referimos ao movimento dialetal que visivelmente se está operando na linguagem nacional."

8 "Haverá impedimento para que eu faça com o verbo reclinar, o que fez Vieira com o verbo inclinar, tão parente e conjunto do outro? A diversidade do prefixo altera de tal forma a palavra que lhe muda a natureza? Ou será que nós brasileiros só temos o direito de cunhar palavras tiradas do *tupi*, como *cuia* e *tiquara*, sendo-nos vedado tocar na arca santa do classicismo?"

Appeal to some "general rule", the lack of a clear theoretical orientation, the erratic way of assembling neologisms, the contradictory attempts to validate the linguistic innovations in the classics, or in Latin etymologies, all characterized the arguments of those who argued for the specificity of BP. As a result, they did not have reliable data to support their opinions and, more importantly, could not draw an objective picture of the changes that the Portuguese language was undergoing in Brazil.

The backdrop of the debates was, on the one hand, the evolutionist climate of the time; on the other, the claim for the autonomy of Brazilian culture due to the ethnical (and linguistic) differences of the peoples who moulded it. Just as the Latin language had changed and transformed itself into the Romance languages, European Portuguese was changing in Brazil. At the end of this process, which was brought about unconsciously by the wider population, generally with no formal education, it would be transformed into a Brazilian language, significantly different from Portugal, and richer. It was possible to attest the "superiority" of BP by considering the various loanwords from African and the native languages. Incidentally, the justification for these neologisms was, according to Pinto (1978, XIX), a vital argument in the definition of the Brazilian language: as pointed out by Macedo Soares in the Prologue to his 1889 *Diccionario Brazileiro da Língua Portuguesa*: *Já é tempo dos brasileiros escreverem como se falla no Brazil, e não como se escreve em Portugal* 'It is time for the Brazilians to write as one speaks in Brazil, and not as one writes in Portugal'.

Concerning African and indigenous speakers, they were certainly bilingual and transmitted their mixed variety to their children, a pre-condition for the formation of a creole language, as pointed out for the first time by the Portuguese philologist Francisco Adolpho Coelho (Coelho 1881). The emphasis on the study of the lexicon perhaps hampered the development of the hypothesis of a creole origin for BP at the time. The notion would be revisited (and rejected) in grammatical terms later, in the 20th century, with theoretical support and contemporary data from Brazilian scholars, including Serafim da Silva Neto (1917–1960) and Fernando Tarallo (1951–1992), (cf. contemporary anthologies like Roberts/Kato 1993; Naro/Scherre 2007; Fiorin/Petter 2008), among many others.

Unlike romantic writers, like Alencar, those who argued in favor of the unity of the Portuguese language, did not value popular usage; on the contrary, according to them, writers should purge literature of popular expressions and not endorse their use. Rui Barbosa (1849–1923), Alfredo d'Escragnolle Taunay (1843–1899), Euclides da Cunha (1866–1909), for example, relying on João Ribeiro's, Carneiro Ribeiro's and Mário Barreto's grammatical theories, realized such purist ideals in literature and rejected the concept of a Brazilian dialect (Pinto 1981).

4.2 Grammar

Research programs do not emerge, as we know, in an intellectual vacuum. The emergence of Portuguese language studies in Brazil is inevitably linked to the political and intellectual climate of 19th and 20th century Europe. As a consequence, the shape that these studies took in Brazil inevitably reflects the way Portugal received German, French and Anglo-Saxon academic literature, as well as the place that the Portuguese language studies occupied in the Romance world. In this sense, one cannot exclude the influence on the formation of Brazilian linguistic thought, be it voluntary or not, by Aniceto Gonçalves dos Reis Vianna (1840–1919), Francisco Adolpho Coelho (1847–1919), Epiphanio Dias (1841–1916), D. Carolina Michaëlis de Vasconcelos (1851–1926) and José Leite de Vasconcelos (1858–1941), among others.

At least for those who followed Adolpho Coelho's lessons, it was clear at the end of the 19th century, and the beginning of the 20th, that grammar was not logic and that philology was not metaphysics. This lesson was dutifully learned, at least in theory, by those Brazilian philologists who recognized in Júlio Ribeiro an innovator of grammatical studies in Brazil. In the words of the author of *Grammatica Portuguesa*:

> 'The facts of any language can only be fully elucidated by the historical comparative study of the grammar of this language. Metaphysical explanations, more or less subtle, more or less ingenious, never satisfy. [...] And I do not introduce myself as someone showing novelties: I merely follow the steps of Mr. C. Waldbach and Adolpho Coelho, of Diez and Bopp, of all the masters of philology and linguistics' (Ribeiro [2]1913, [1]1881, 331).[9]

In the new order, the empirical study of linguistic usage (even if that one subscribed by the "great authors") should precede theory; and the explanation of linguistic forms and functions should not be derived from logical constraints, but from their historical causes. It was the radical subversion, in relation to the previous generation, of general principles that dominated linguistic analysis.

In Brazilian intellectual circles of the end of the 19th century and of the first half of the 20th century, individuals of different orientations still operated under the vast umbrella term philology, such as Francisco Sotero dos Reis (1800–1871), Ernesto Carneiro Ribeiro (1839–1920), Júlio Ribeiro (1845–1890), Fausto Barreto (1852–1915), Manoel Pacheco da Silva Jr. (1842–1899), Mário Barreto (1879–1931), Alfredo Gomes (1859–1924), Eduardo Carlos Pereira (1855–1923), Maximino Maciel (1865–1923), João Ribeiro (1860–1934), Manuel Said Ali (1861–1953), Amadeu Amaral (1875–1929), Otoniel Mota

9 "Os factos de uma lingua qualquer só podem ser cabalmente elucidados pelo estudo historico comparativo da grammatica dessa lingua. As explicações metaphysicas, mais ou menos subtis, mais ou menos engenhosas, nunca satisfazem. [...] E não me apresento como exhibindo novidades: sigo apenas os passos dos Srs. C. Waldbach e Adolpho Coelho, de Diez e Bopp, de todos os mestres de philologia e linguistica."

(1878–1951), José Oiticica (1882–1957), Sousa da Silveira (1883–1967) and Antenor Nascentes (1886–1966) (Coelho 1998, 81).

All of them were self-taught philologists, having trained in other disciplines, such as engineering, law, or medicine. Like a large number of other professionals, their ideas circulated in newspapers, magazines, and recently created journals like *Revista Brazileira* (Rio de Janeiro, Laemmert, 1895–1898), *Revista de Língua Portuguesa. Archivo de Estudos Relativos ao Idioma e Literatura Nacionais* (1919–1935, dir. by Laudelino Freire), and *Revista de Filologia Portuguesa* (São Paulo, Nova Era, 1924–1925, dir. by Mário Barreto) (Christino 2001; Altman 2014).

The grammatical manuals within the new framework proposed the phonetic, morphological and syntactic history of the Portuguese language, and the study of its etymology. This does not mean, however, that the reception of historical-comparative grammar was always coherent: the issue of the pronoun placement in the Portuguese sentence is a good example. Although claiming the need to deal with usage data, the examples given tended to be drawn from Portuguese literature, or, at most, from the author's own repertoire (Praxedes 2008). This topic provoked hot debate among European and Brazilian grammarians at the end of the 19th century, and illustrates how impressionistic and nationalistic their argumentation still was, with the possible exception of Said Ali (1895), who stressed the importance of prosodic factors in the discussion of the problem.

Regarding the differences between European and Brazilian Portuguese, the majority of the examples offered by the authors concerned syntax; lexicon comes in second place, with phonetics and morphology in third and final place, respectively (Coelho et al. 2014, 121). Besides word order within the sentences (including the use of the pronouns), the main syntactic differences discussed had to do with agreement, government, the existential use of the verb *ter* 'to have', the construction of relative clauses, negative sentences, and the use of the article (cf. Coelho et al. 2014, 123ss.). On the characteristics of present-day BP syntax, see ↗14 Syntax.

If neologisms were judged positively, in that they enriched the language, the syntactic particularities of the Brazilian variety were considered incorrect, seen as errors to be avoided.

Observe Ribeiro ([2]1913, [1]1881, 230):

> 'Every word that serves as an objective complement to a verb is put in the accusative case. As in Portuguese, nouns are not inflected, the application of this rule only becomes patent when the objective complement is represented by a substantive pronoun, ex: *I see him—You love me very much*. To put into the nominative case the substantive pronoun that serves as an objective complement to a verb is an every-day error in Brazil, even among learned people: one hears at every step the incorrect phrases: *I see he—Wait I*.'[10]

10 "Toda a palavra que serve de objecto a um verbo põe-se em relação objectiva. Como em Portuguez não se declinam substantivos, a applicação desta regra só se torna patente quando o objecto é repre-

Be that as it may, this generation of Brazilian philologists, oriented to the Portuguese historical dimension, was not overly concerned with handling contemporary oral language data. So much so that it was in fact a man of letters, not a philologist, who would present a new way of investigating Brazilian dialect formation. Amadeu Amaral inaugurated what would be considered by this same generation a new research program to follow, see also ↗9 Dialectology and linguistic geography. The goal was to achieve, through the accumulation of partial monographs, a portrait of the Brazilian dialect. In his words:

> 'It would be desirable that many impartial observers, patient and methodical dedicated themselves to collect elements from each one of these regions, limiting themselves strictly to the familiar field and banning completely all that was hypothetical, uncertain, not personally checked' (Amaral [3]1976, [1]1920, 43).[11]

This clearly underlines the need felt by the author to pay more attention to oral language. Between 1930 and 1940, dozens of monographs were presented at the *Colégio Pedro II* describing Brazilian dialects from North to South, most notably vocabulary (cf. Dietrich 1980). Only Nascentes (1922) followed Amaral's full model, also describing grammatical aspects of the *carioca* dialect (i.e., the variety from Rio de Janeiro). The study of the Brazilian rural variety contributed to a justification of the divergent interpretations of the substrata of BP. The larger field of philology incorporated Brazilian dialectology as a program of data collection of the regional varieties of BP. So much so that it was through the initiative of the celebrated Brazilian philologist Serafim da Silva Neto (1917–1960) that Brazilian dialectology came to enjoy, in the 1950s, a strong institutional position.

5 Philologists and linguists of the 20th century

The beginning of scientific language studies, as we know them today, is often associated with the creation of the first 'Colleges of Arts' (*Faculdades de Filosofia*), in São Paulo (1934) and Rio de Janeiro (1935). The establishment of these colleges marked the end of self-education in linguistic matters and symbolized, for the following generations, the beginning of the era of professional careers in the study of language and literature.

---

sentado por um pronome substantivo, ex.: *Eu o vejo—Queres-me muito*. Pôr em relação subjectiva o pronome substantivo que serve de objecto a um verbo, é erro comezinho no Brazil, até mesmo entre os doutos: ouvem-se a cada passo as locuções incorrectas: *Eu vi ele—Espere eu*."

11 "Seria de se desejar que muitos observadores imparciais, pacientes e metódicos se dedicassem a recolher elementos de cada uma dessas regiões, limitando-se estritamente ao terreno conhecido e banindo por completo tudo quanto fosse hipotético, incerto, não verificado pessoalmente."

Scholars like Ernesto de Faria (1906–1962), Joaquim Mattoso Câmara Jr. (1904–1970), Silvio Edmundo Elia (1913–1998), Serafim Pereira da Silva Neto, Theodoro Henrique Maurer Jr. (1906–1979), Gladstone Chaves de Melo (1917–2001), Isaac Nicolau Salum (1913–1993), Francisco da Silveira Bueno (1898–1989) and Celso Ferreira da Cunha (1917–1989) occupied the main professorial chairs in the country, founded the first specialized linguistic research centres, and enlarged and dominated monographic and periodical publication.

The first linguistics course in Brazil, given by Mattoso Câmara Jr. between 1938 and 1939 at the College of Arts, Rio de Janeiro, presented to the scholarly community of the time a framework partially different of the one adopted by Brazilian philologists: Sapir, Jespersen, Meillet, Vendryès and Saussure were the most frequently quoted authors in Mattoso Câmara Jr.'s lectures (França 1995), which were published for the first time in *Revista de Cultura Vozes* between 1939 and 1940 under the title *Lições de Linguística Geral* 'Lessons in General Linguistics', and which were disseminated from 1941 onwards by successive editions of his *Princípios* (cf. Câmara Jr. [1]1941). For this work, Mattoso would be distinguished as the "father of modern linguistics" in Brazil. The book symbolizes the transition between a long-standing tradition in historical and comparative studies within Portuguese philology, dominant in Brazilian academic circles since the end of the 19th century, and an emergent synchronic and autonomous orientation in the study of (the Portuguese) language.

As a matter of fact, as early as 1939 Mattoso Câmara Jr. seemed convinced of the originality and value of the approach which he called "static" linguistics—in opposition to "evolutive" linguistics—whose main tasks, according to him, were the study of the phonetic system and its features, the study of the morphological system, the study of the phrase, and the unity of speech, as well as the study of general topics like vocalic alternation and language classification (Câmara Jr. 1939, 101).

Mattoso Câmara Jr.'s structural(ist) analysis in the 1940s or even the 1950s, however, were still far from representing the mainstream at the time; the philological approach to the study of Portuguese and its dialectological varieties, as well as in the study of other Brazilian languages, were still the dominant ones, at least as far as academic circles were concerned. Brazilian scholars' attention at this moment had turned in a different direction: the "dialectological crusade" proposed by Silva Neto for the development of a dialectological atlas of Brazil (cf. the creation, in 1953, of a *Center of Dialectological Studies* at the National Museum, in Rio de Janeiro and of a new journal, *Revista Brasileira de Filologia* vol. 1–5, 1955–1959/1960).

Despite this, the emergence of an autonomous field of teaching and research denominated linguistics, officially created in Brazil in 1962 by a Federal decree, is linked to the leadership of various Brazilian scholars of the period, some of them linguists *avant la lettre*, who favored the diffusion and reception of those frameworks which had been somewhat in evidence in Europe, but also in United States, between the 1930s and the 1960s. We can say, in retrospect, that the establishment of "modern" linguistics in Brazil is linked to Mattoso Câmara Jr.'s structural analysis of Portuguese

and also to Aryon Dall'igna Rodrigues' (1925–2014) structural analysis of Brazilian indigenous languages.

According to Coseriu (1976), European structuralism of Prague School had been introduced into South America during the 1940s and had spread across the continent during the 1950s. So-called American structuralism would become known even later. From the second edition of Mattoso Câmaras's *Princípios* (1954) on, however, we can observe more and more references to North American linguists. Authors like Bloomfield, Boas, Hall, Nida, Whorf slowly became part of Brazilian linguistic literature. An external fact contributed to the entrance of the American Schools of Linguistics into the Brazilian scene: Mattoso's celebrated trip to United States from September 1943 until April 1944, sponsored by the Rockefeller Foundation. Certainly, it was during this visit that Mattoso became acquainted not only with Bloomfield's descriptivism, but also (and mainly) with Prague School structuralism, through Roman Jakobson and the Linguistic Circle of New York.

The explicit refusal to embrace the descriptivist framework in a philological fortress like the Faculty of Arts of the University of São Paulo is a good illustration of the values which the "two structuralisms", the European and the North American, were assuming among Brazilian scholars during the 1960s. Theodoro Henrique Maurer Jr. (1906–1979), Indo-Europeanist, religious leader and Professor of Romance Philology, had been at Yale between 1945 and 1946, where he followed Bloomfield's lectures. Unlike Mattoso, however, the only residue of his American experience can be found in a preface to a bulletin of the faculty which published a work about a native language (Maurer 1952). Except for this text, in which Maurer reviews Bloomfield linguistics in a positive light, his option was always for the philological approach, meaning in this context the historical approach (cf. Altman 1999).

On the one hand, if the reception of "American structuralism" had been negative among the philologists dedicated to the study of Portuguese, it seems, on the other, to have been more positive among the anthropologists at the National Museum in Rio de Janeiro, who began to work on the description of native languages under the supervision of the linguists of the Summer Institute of Linguistics (SIL), which inaugurated its activities in Brazil in 1956 (Magalhães 1981, 755). The agreements between SIL and the National Museum favored the structuralist approach in the study of Brazilian native languages, and at the same time contributed to the establishment of an autonomous space for the discipline of Amerindian linguistics (Barros 1993). Indeed, in 1958, Mattoso created, at the National Museum in Rio de Janeiro, a *Linguistics Sector* alongside Silva Neto's *Dialectology Sector*, although completely autonomous from it.

The controversial activities of the SIL grew in importance in the 1960s and 1970s in what concerned the task of describing Brazilian native languages, as well as offering elementary education to native peoples and, quite often, linguistic analysis training for linguists. In this connection, besides the National Museum of Rio de Janeiro, SIL developed, until the late 1970s, several exchange programs with the University of Brasília, the Federal University Federal of Rio de Janeiro, the Fundação Nacional do

Índio (Funai), the Museu Emilio Goeldi (Belém) and the State University of Campinas (São Paulo), among others. A great part of linguistic studies in Brazilian native languages of the 1960s, and even the 1970s, was done within the Tagmemics framework, along with less numerous studies based on indigenous toponymy and onomastics (cf. the literature presented in Magalhães 1974; 1975).

It was too late for the development of a structuralist research program in Brazil, however. By the end of the 1960s, studies in BP norm, both written and spoken, as well as studies in native languages, adopted a more "theoretical" conception of language science and condemned our newly-born structuralism to the status of an anachronism even before it had truly taken hold (Altman 1994).

6 Contemporary BP linguistics

Ever since its official creation by decree in the 1960s, Brazilian linguistics has witnessed a process of immense expansion. Some of the most important achievements are the establishment of several important linguistic institutions, the success in refuting prejudices about BP, i.e., establishing BP as a topic of investigation in its own right and developing more adequate educational materials, the thorough description of spoken BP, and the investigation of its history.

Contact and exchange with the international linguistic community also increased continuously. This exchange brought all the different schools of modern linguistics to the country, where research groups and networks had to be formed in and around the profiles of researchers, language and linguistics departments, and subdisciplines. However, only few decades later, the "flow of information" started to become less and less unidirectional, so that nowadays BP has become one of the best-described spoken Romance vernaculars, receiving considerable attention around the world, indeed, beyond Romance linguistics.

6.1 Institutionalization and emancipation

As already mentioned in Section 5, linguistics was established in Brazil in 1962 as an obligatory discipline in university Languages and Literature courses. In the years immediately before that, courses featuring the label *linguistics* had been taught only in three Brazilian universities: Indo-European and Romance Linguistics at the University of São Paulo (by Theodoro Henrique Maurer Jr.), Descriptive Linguistics at the University of Rio de Janeiro (by Joaquim Mattoso Câmara Jr.), and General Linguistics at the University of Paraná (by Raul Farâni Mansur Guérios (1907–1987)). It is important to bear in mind, however, that notable researchers who considered themselves primarily as philologists or dialectologists were also aware of contemporary issues in linguistics and participated in the discussion of these.

Especially from the late 1960s and 1970s, several initiatives were taken to improve linguistics teaching and research. Several linguistic societies were founded, such as the *Associação Brasileira de Linguística* ('Brazilian Association of Linguistics', ABRALIN), the *Grupo de Estudos Linguísticos do Estado de São Paulo* ('São Paulo Group of Linguistics', GEL), both in 1969, and the *Associação Nacional de Pós graduação e Pesquisa em Letras e Linguística* ('Graduate Association of Linguistics and Literature', ANPOLL, 1984), among others. Furthermore, linguistic journals now emerged, among them *Alfa* (1962), *Cadernos de Estudos Linguísticos* (1978), *D.E.L.T.A.* (1984), *Filologia e Linguística Portuguesa* (1997), and many others. Last but not least, collective projects of research were launched, such as the *Projeto da Norma Urbana Culta* ('Project of Urban Educated Norm', NURC, 1967), the project *Gramática do Português Falado* ('Description of Spoken Brazilian Portuguese Project', 1988), *Projeto para a História do Português Brasileiro* ('Towards a History of Brazilian Portuguese', PHPB, 1997), and *Programa de Estudos sobre o Uso da Língua* ('Programme of studies about language use', PEUL), among many others. As a consequence of this, a new perception of BP took place, emancipating the study of the vernacular ever more from the traditional comparison to European Portuguese and continuously putting aside also prejudices about its inferiority with regards to the latter. From the several collective projects that took spoken BP as their object of study, the following section will focus on two of the most important ones, the NURC project and the project on Spoken Portuguese Grammar.

6.2 The description of the spoken vernacular

6.2.1 The NURC project

In 1969, a group of Brazilian linguists established a project for the documentation and description of BP as spoken by Brazilians holding a university degree. This project echoed a Hispano-American one, which had been proposed in 1967 by Juan M. Lope Blanch (1927–2002), at that time a professor at the Colegio de México. The Brazilian edition, which would be known as the NURC Project, was initiated by Nelson Rossi (1927–2014), a professor at the Federal University of Bahia. He chose the cities to be studied, four of them founded in the 16th century (Recife, Salvador, Rio de Janeiro and São Paulo), and one founded in the 18th century (Porto Alegre). The organization of the documentation teams was carried out by local coordinators, who also obtained the necessary funding. The NURC corpus was compiled between 1970 and 1977. Three activities were undertaken by the researchers: recording of interviews, the transcription of the recorded materials, and their lexical and morpho-syntactic analysis. The interviews had a guiding questionnaire and followed a sociolinguistic methodology. Different types of oral production were recorded: Formal speech (for instance teaching situations, conference talks), interviews between an informant and a researcher, and

conversations between two informants. Transcriptions were provided in a semi-orthographic style. These were not intended as final results for analysis but rather for making the content of the interviews more easily accessible and for having an overview of what kind of data could be expected. For more in-depth linguistic analysis of spoken language, the recordings have to be accessed directly. In order to coordinate the ongoing activities, fifteen meetings were held in the cities involved, the last of which took place in 1989. Detailed information on this process can be found in Castilho (1970; 1984; 1990).

The analyses of these materials were launched from 1978 onwards, initially taking as its starting point the Guiding Questionnaire of the project, which was divided into three parts: Phonetics and phonology, lexicon, and morphosyntax. Summing up, for the five aforementioned state capitals, the project produced a total corpus of 1,870 interviews, recorded with 2,356 informants, totalling 1,570 hours, the most extensive documentation of Portuguese spoken in Brazil. Samples of these materials from all the cities involved were published: Castilho/Preti (1986; 1987); Preti/Urbano (1989); Callou (1992); Callou/Lopes (1993; 1994); Motta/Rollemberg (1994; 2006); Sá et al. (1996; 2005); Hilgert (1997).[12] For a complete bibliography until 1990, see Castilho (1990a, 141–185).

While the compilation of the NURC corpus and its processing can be considered a great achievement, there were also several difficulties that had to be dealt with during data collection and analysis. For instance, the Guiding Questionnaire turned out not to be ideal for the registration of many spoken language phenomena. Lacking an explicit hypothesis about orality at that time, the questionnaire that the interviewers followed was constructed by combining the perspective of a sort of moderate structuralism with elements taken from traditional grammars, which are based on the description of the written language. Looking back, the project might usefully have paid more attention to contemporary developments of linguistic theories dealing with the interface between syntax and discourse (for a detailed analysis of such issues in the early works of the NURC project, see Castilho 1984; 1990). However, the obstacles encountered also led to new directions in the exploration of the NURC corpus, such as conversation analysis (Marcuschi 1983; 1991; 1997; Preti/Urbano 1990; Preti 1993; 1997; 1998) and the relationship between spoken and written language (Preti 2000; 2002; 2005; 2006). The corpus was also analyzed with respect to particular phenomena of BP grammar, for instance in Dias de Moraes (1987), Marcuschi (1991), Castilho (1989), among many others. The huge amount of collected data and the insights gained from the first analyses paved the way for a new large-scale project which would underline the legacy of the NURC project and keep its spirit alive.

12 Samples of this huge corpus can be found on the sites of several of the subprojects: https://fale.ufal.br/projeto/nurcdigital/, https://nurcrj.letras.ufrj.br/, http://nurc.fflch.usp.br/, last accessed 13.02.2022.

In 2019, the NURC project celebrated its 50th anniversary (Oliveira Jr. 2019). While the collected data are still used for current investigations, new issues have gained more prominence, such as digitalization and long-term archiving of this linguistic treasure. The data have had, and continue to have, an inestimable impact on how BP is seen and understood not only in the (inter)national linguistic community, but also, perhaps more importantly, by its speakers.

6.2.2 The 'Grammar of Spoken Brazilian Portuguese' project

Not least because of the previously mentioned challenges in the original grammatical analyses based on the NURC data, a new project was developed, aiming at collectively preparing a usage grammar based on its own corpus. First aspects of this new plan were discussed at a meeting in October 1987 at the Department of Linguistics of the State University of Campinas, and subsequently in the first seminar of what in 1988 was baptized *Projeto Gramática do Português Falado* ('Grammar of spoken Portuguese Project', PGPF). Thirty-eight researchers from twelve Brazilian universities joined PGPF. Milton do Nascimento acted as an academic advisor for the project, responsible for sketching the theoretical directions to be taken by researchers. At the same time, PGPF started to exchange ideas with international experts in spoken language description and with colleagues that had experience in the preparation of collective grammars, such as Claire Blanche-Benveniste (Université d'Aix-en-Provence), Randolph Quirk (University College London), Giampaolo Salvi (University of Budapest), Maria Fernanda Bacelar do Nascimento (Centre for Linguistic Studies, University of Lisbon), Françoise Gadet (Université Paris-Nanterre), Brigitte Schlieben-Lange (University of Tübingen), among others. The following strategic decisions were taken, based on these preparatory steps:

(1) The specific objective of PGPF is to "prepare a reference grammar of Portuguese spoken in Brazil, describing its phonological, morphological, syntactic, textual and pragmatic levels" (Castilho 1990b, 9–27).
(2) Given the enormous scope of the project, coexistence between different theoretical positions, mainly formalists and functionalists, was encouraged, since one single theory would not be able to account for all the issues of a usage-based grammar the project was expected to have to deal with.
(3) The following methodology was established: (i) use of three of the interviews from the NURC project per city, i.e., 15 in total; (ii) distribution of researchers in six working groups (WG), each one with its own theoretical proposal and schedule of activities. The following WGs took a formalist perspective: phonetics and phonology (coordinated by Bernadette Maria Marques Abaurre), derivational morphology (coordinated by Margarida Basilio and later on by Ieda Maria Alves), syntax of grammatical relations (coordinated by Mary Aizawa Kato). Further WGs followed a functionalist perspective: inflectional morphology (coordinated by Angela Ceci-

lia Souza Rodrigues), syntax of word classes (coordinated by Rodolfo Ilari and later on by Maria Helena de Moura Neves) and interactive textual organization (coordinated by Ingedore Grünfeld Villaça Koch).

Eight plenum seminars were held between 1988 to 1998 to discuss the results produced by the WGs. Many of these results were published in a series of eight volumes (Castilho 1990b; 1993; Ilari 1992; Castilho/Basílio 1996; Kato 1996; Koch 1996; Neves 1999; Abaurre/Rodrigues 2002). More general, theoretical issues were discussed in Nascimento (1993); Castilho (1997a; 1997b; 2009).

From 1998 onwards, condensed results, together with further analyses, were consolidated and published as a collectively written grammar, in a series of seven volumes: *A construção do texto falado* 'The construction of spoken text' (Jubran [2]2015), *A construção da sentença* 'Construction of the sentence' (Kato/Nascimento 2015), *Palavras de classe aberta* 'Open class words' (Ilari 2014), *Palavras de classe fechada* 'Closed class words' (Ilari 2015), *A construção das orações complexas* 'Construction of complex sentences' (Neves 2016), *A construção morfológica da palavra* 'Morphological construction of the word' (Alves/Rodrigues 2015), *A construção fonológica da palavra* 'Phonological construction of words' (Abaurre 2013). The PGPF, thus, provided a careful description of spoken BP as practiced by educated people in the country. It also represents a strong coalition of Brazilian linguists with solid theoretical preparation and a relish for dealing with empirical data. It would be impossible to present all the accomplishments of PGPF in this overview. However, many of the synchronic chapters of this manual draw heavily on them. It might, then, be enough to point out that no other Romance language has had the variety spoken by its population with higher education as fully described as BP.

6.3 Grammars written by linguists

One of the main characteristics of the first Brazilian scholars who, in the sixties and the seventies, saw themselves as linguists, was the rejection of the academical philological establishment and traditional normative grammar. They attacked what they considered to be inconsistencies in the normative approach, or, even worse, they questioned the usefulness of the philological framework as a whole. These "new linguists" were happy to describe such inconsistencies in normative grammar and the perceived futility of philology, whose main concern in those days was historical grammar. Over time, several Brazilian linguists began to write their own grammars, trying to overcome what had been identified as flaws in traditional grammaticography. Different approaches were taken in search for an adequate way to achieve these goals, resulting in a collection of grammars of quite different profiles.

The first grammar of this type was published by Mário Perini in 1995: *Gramática descritiva do português*. It takes great care in explaining and defending the new ap-

proach. Several dozen pages are dedicated to this theme, which is indeed the main topic in the preface, introduction, and first chapters. One of the innovations here is that the intuitions of native speakers are introduced as a valid data source in addition to examples from the classic authors. The grammar also advises its readers that the language of the text itself will deviate from what is traditionally found in grammars and that it will approximate to the modern standard in Brazilian technical and journalistic prose. Prescriptive norm, as found in traditional grammars, is considered to be "ficticious". Perini updated this work twice, first in 2010, *Gramática do português brasileiro*, this time including reference to Brazil in the title and emphasizing the description of the spoken language more prominently; and a second, considerably extended edition in 2016, *Gramática descritiva do português brasileiro*. Furthermore, in 2002, he published a grammar for non-native speakers in English, *Portuguese: a reference grammar*, which also deals with BP.

In 2000, Maria Helena Moura Neves published her monumental *Gramática de usos do português*, a reference grammar of modern BP based on its own text corpus of 70 million tokens with a wealth of examples representing current usage. This grammar is also innovative in consistently exploring new ways of structuring and presenting the linguistic facts – not based on the traditional parts of speech but on more general categories and processes, such as (i) predicate formation, (ii) reference, (iii) quantification and (iv) junction. This work was considerably revised and extended into the even more monumental *A gramática do português revelada em textos*, published in 2018, which also contains spoken language texts and thus explores an even more diverse domain of language use.

The grammars described above are clearly (and sometimes explicitly) dedicated to rather academic contexts of use. For a less specialized readership, or even for use in secondary schools, there are several works that provide normative guidance and exercises on the one hand, but which also seek to connect readers more directly with the results of current linguistic research, especially concerning linguistic variation. Among these are *Gramática Houaiss da língua portuguesa* by José Carlos de Azeredo (2008), Ataliba Teixeira de Castilho's and Vanda Maria Elias' *Pequena gramática do português brasileiro* (2012), and Lorenzo Vitral's *Gramática inteligente do português do Brasil* (2017).

Finally, two recent grammars have to be mentioned that are not only radically innovative in their description of the language, but which also translate their ideology into a research and teaching program and intertwine both into a further type of "grammar". Ataliba Teixeira de Castilho (2010) and his *Nova gramática do português brasileiro*, in addition to describing the grammar of spoken BP, also provides an introduction to linguistics with ample advice for students in approaching linguistic problems; it also introduces a new perspective on linguistic research within the framework of the theory of language as a complex system. In his preface to the book, Rodolfo Ilari qualifies the title of the book as the "least foreseeable one" or even "paradoxical" for its content and observes that it is full of conscious transgressions and

"provocations" with respect to traditional grammar. Marcos Bagno's *Gramática pedagógica do português brasileiro* (2011) is a monumental work, one which as well as describing contemporary BP (including spoken langue) in great detail is strongly concerned with issues of teaching, as the title indicates. Not only does it argue for a radically new approach of teaching grammar in order to reduce linguistic prejudices towards spoken BP structures and in order to make teaching of grammar more functional, it also contains regular alerts in highlighted boxes as to what NOT to teach at school or in language classrooms, thus directly attacking traditional grammar in this domain. With regard to the language of the text, it obviously also privileges forms that are in current use over those prescribed forms from normative grammar which are largely out of use in Brazil; indeed, it goes beyond the other grammars, such as Perini's, in freely using structures that are highly stigmatized by the grammatical tradition. While the merits of these new grammars are beyond any doubt, they have also received critical reactions, among other things for the—admittedly knotty—attempt to provide a unified description of written and spoken language. Furthermore, traditional grammarians did not willingly accept the role of the scapegoat for the existence of language prejudice in the country, and saw their position as somewhat misrepresented (cf. Cavaliere 2012).

7 The 'History of Brazilian Portuguese project' (PHPB)

In order to account for the modern language and its variation, one fundamental path of explanation is the historical one. How did the situation as described in all its complexity come about? Most research in this respect for Brazil happened and is still happening in the context of the *Projeto Para a História do Português Brasileiro* (PHPB). PHPB was launched in 1997 at the University of São Paulo, initially to unveil the formation of *paulista* BP, since lusitanization of the Brazilian territory took place first in São Vicente, located on the *paulista* shores, from 1532 on. However, very soon, the idea of the project spread to 14 Brazilian regions, gathering around it about 200 researchers into PHPB (Castilho 1998c).

Among the main motivations that drove PHPB were the following three: first, after decades of dedicated focus on the contemporary language, it was time to strengthen once more the tradition of historical linguistics at Brazilian universities. Second, the diachronic roots of the countless phenomena described by NURC and PGPF were largely unknown and had to be revealed (Moraes de Castilho 2001; 2013, among others). Widely diverging speculations about language-internal tendencies as drifts, as opposed to developments out of linguistic contact towards putative creolization scenarios, were in the air, although with little empirical foundation. And third, the linguistic community could take advantage of the encouraging experiences accrued by the pre-

vious collective projects. For PHPB, two especially relevant prior projects were PROHPOR, the 'History of Portuguese Programme', developed in the eighties at the Federal University of Bahia by Rosa Virginia Mattos e Silva (1940–2012), and the "marriage" of Labovian variationism with the theory of Principles and Parameters, inaugurated at the State University of Campinas by Mary Kato and Fernando Tarallo (Roberts/Kato 1993). Thus, since its origin, PHPB stimulated the acquaintanceship of contrary theoretical orientations, bringing together sociolinguists, generativists, functionalists and cognitivists. This also included the invitation of foreign specialists, many of them from Germany (Brigitte Schlieben-Lange, Volker Noll, Wulf Oesterreicher, Johannes Kabatek, Konstanze Jungbluth, Eberhard Gärtner, Roland Schmidt-Riese, Uli Reich) and from Portugal (Ivo Castro, Augusto Soares da Silva, Clarinda Maia), who also contributed new theoretical perspectives.

The main goals of PHPB were to organize a diachronic corpus, to investigate the social history of BP, and to describe the grammatical change in the language, its discourse traditions, and the history of the texts. A proper publication series was founded, in which most of the papers that were prepared for the (thus far) nine national meetings of the project were published: I, São Paulo, 1997: Castilho (1998c); II, São José dos Campos, 1997: Mattos e Silva (2001); III, Campinas, 1999: Alkmim (2002); IV, Teresópolis, 2001: Callou/Duarte (2002); V, Ouro Preto, 2002: Ramos/Alkmin (2007); VI, Ilha de Itaparica, 2004: Lobo et al. (2006); VII, Londrina, 2007: Aguilera (2009); VIII, João Pessoa, 2010: Hora/Rosa (2010); IX, Maceió, 2013 (unpublished). The eighth meeting eventually decided that the time had come to condense and consolidate the results so far obtained, giving birth to a new series on the history of BP.[13] The diachronic chapters in this manual are mostly based on the findings of PHPB.

While this major series is currently publishing consolidated results about the history of BP, research continues on all levels. By searching systematically for older texts that reflect BP phenomena, researchers are digging deeper and deeper into the past and the corpus of PHPB keeps growing, see ↗0 Introduction: the state of the art in Brazilian Portuguese linguistics for details. The corpus comprises newspaper ads, personal letters, administrative letters, theatre plays, wills and testaments, proceedings of city councils, law suits, historical memoires, news, chronicles, spoken BP, covering a period from the 16th century until now. Text types that reflect phenomena of spoken

13 The volumes cover the most important domains of linguistic description: general historical overview (Castilho 2018), presentation of the corpus (Castilho 2019a), sound change (Hora/Battisti/Monaretto 2019), syntactic changes in word classes from a functional perspective (Lopes 2018), constructional syntactic changes from a functional perspective (Castilho 2019b), syntactic change from a generative perspective (Cyrino/Torres Morais 2018), discourse traditions (Andrade/Gomes 2018), diachronic processes of textual construction (Penhavel/Cintra in print), semantic change (Ilari/Basso 2020), the social history of BP (Callou/Lobo 2020), the relationship between linguistic and social history (Ramos/Oliveira 2021), and the history of the lexicon (Aguilera/Altino in prep.). A complete bibliography of the more than 700 publications in the context of PHPB can be found in Castilho (2019c).

language (e.g., personal letters and theatre plays) are of paramount importance for documenting the existence and origin of modern Brazilian structures, as are texts written by speakers less exposed to literary language and European norms. The early 20th century as well as the 19th century had been the main focus up until now, and the documentation here is quite solid. For the 18th century and before, there is far less documentation, especially for spoken language phenomena. One major task for the future will be to provide a unified digital infrastructure with a search interface for this project, a dream which has yet to come to fruition.

8 Outlook and new issues

The generation now about to leave the scene has worked hard to set up linguistics in the country, founding graduate programs and specialized journals, creating professional associations, managing congresses and seminars, launching collective research projects, and focusing on central themes of Brazilian linguistic identity. During the "founding decades", the usual research attitude of linguists was to look for a renowned foreign linguist, translate his or her texts, adhering (maybe uncritically) to their theory, and analysing data according to it. No other way was available. Yet if this routine continues today, Brazilian linguistics risks becoming irrelevant. This raises the question about the future orientation of the field, and in the following section, some thoughts on this topic are offered.

First of all, linguists should maintain and strengthen the collective way of doing science. Brazilian linguists did so in a very fruitful way, as described above, as several co-authored publications of authors from different theoretical orientation (Galves/Lopes, Galves/Oliveira, Longhin/Paixão de Souza) in this manual show. In addition, it would be profitable to think about new ways of deriving generalizations based on the extensive empirical data now available. First attempts in this direction are already visible, pointing towards at least two broad and somewhat distinct means of achieving such a goal. One approach would be to retain classical scientific epistemology. This is by far the predominant approach and can be found throughout the subsequent chapters of this manual. Given the solid empirical foundation built so far, it is in a good position to challenge Eurocentric theories, as witnessed repeatedly in the near past, and to contribute important insights from a Romance language that has developed a very distinct profile in recent times.

The other perspective would be to approach language as a complex system, from a dedicatedly non-deterministic perspective. First references to this second approach can be found in Franchi (1977), where he assumes language to be a constitutive activity, as "a job that gives form to the variable content of our experiences". Language is "a work in construction", it is a process, not a mere instrument of communication—which recalls the Humboldtian assumption of language as *enérgeia*. The assumption of language as a complex system fits better when we deal with language as an activity,

as a form of action. To deepen this perspective, a new research group has been created at the Catholic University of Minas Gerais, the "Complex Cognition Group": Paiva/Nascimento (2009). Studies on grammaticalization are also tentatively elucidating the notion of language-as-a-process. What remains is to incorporate other processes of language creativity, discarding its current epiphenomenal approach. For further developments of these ideas and their testing, see Castilho (1997a; 1997b; 1998a; 1998b; 2010), Castilho et al. (2019); Módolo (2004), Kewitz (2007), Simões (2007). It goes without saying that a wide range of skills will be required to develop such an agenda into a fully-fledged research program. One single researcher cannot hope to master all of them. As a consequence, collective research will be needed and will have to extend to new domains. A new collective project may be on its way.

Summing up, the task of describing the current dynamics in the evolution of BP and its history (and of relating the two) will continue, contributing to a better understanding of the *Romania Nova*. Brazilian linguists are obviously also expected to deepen our knowledge of the indigenous languages spoken in the country. Recent collective international publications on the typology of argument structure (Gildea/Queixalós 2010), the sound systems of these languages (Avelino/Coler/Wetzels 2015) and their syntax and semantics in the nominal domain (Lima/Rothstein 2020) are just some highlights of contributions to general linguistics which, together with the following chapters of this volume, can give us an idea of the potential of Brazilian (Portuguese) linguistics in the future.

9 References

Abaurre, Maria Bernadete Marques (ed.) (2013), *Gramática do Português Culto Falado no Brasil*, vol. 7: *A construção fonológica da palavra*, São Paulo, Contexto.

Abaurre, Maria Bernadete Marques/Rodrigues, Ângela C. S. (edd.) (2002), *Gramática do Português Falado*, vol. 8: *Novos estudos descritivos*, Campinas, Universidade Estadual de Campinas/Fapesp.

Aguilera, Vanderci de Andrade (ed.) (2009), *Para a História do Português Brasileiro,* vol. 7: *Vozes, veredas, voragens*, 2 tomos, Londrina, Universidade Estadual de Londrina.

Aguilera, Vanderci de Andrade/Altino, Fabiane Cristina (edd.) (in prep.), *História do Português Brasileiro*, vol. 12: *Léxico Histórico do Português Brasileiro*, digital version.

Alencar, José de (1873), *Questão Filológica*, in: Pinto (ed.) (1978), 96–121.

Alkmim, Tânia (ed.) (2002), *Para a História do Português Brasileiro*, vol. 3: *Novos Estudos*, São Paulo, Humanitas/Universidade de São Paulo; Instituto de Estudos da Linguagem/Universidade Estadual de Campinas.

Altman, Cristina (1994), *Trinta anos de linguística brasileira. Movimentos de afirmação e auto-afirmação profissional*, D.E.L.T.A. 10/2, 389–408.

Altman, Cristina (1999), *Between Structure and History: The search for the specificity and the originality of Brazilian linguistic production*, in: Sheila Embleton/John Joseph/Hans-Joseph Niederehe (edd.), *The Emergence of the Modern Language Sciences*, Amsterdam/Philadelphia, Benjamins, 247–259.

Altman, Cristina (2007), *Colonialism, scientific expeditions and linguistic typology in 19th century Brazil*, in: Douglas Kibee (ed.), *History of Linguistics 2005*, Amsterdam/Philadelphia, Benjamins, 212–227.

Altman, Cristina (2014), *O pêndulo de Foucault. Sincronia e Diacronia no Estudo do Português do Brasil*, paper given at the XIII International Conference on the History of the Language Sciences, Vila Real, Portugal, 25–29 August, unpublished.

Alves, Ieda M./Rodrigues, Ângela C. S. (edd.) (2015), *Gramática do Português Culto Falado no Brasil*, vol. 6: *A construção morfológica da palavra*, São Paulo, Contexto.

Amaral, Amadeu ([3]1976, [1]1920), *O dialeto caipira. Gramática – Vocabulário*, São Paulo, Hucitec.

Anchieta, Joseph de ([7]1990), *Arte de grammatica da lingoa mais usada na costa do Brasil*, São Paulo, Loyola; ([1]1595), Coimbra, Antônio Mariz.

Andrade, Maria Lúcia C.V.O./Gomes, Valéria (edd.) (2018), *História do Português Brasileiro*, vol. 7: *Tradições discursivas do Português Brasileiro: constituição e mudança dos gêneros discursivos*, São Paulo, Contexto/Facepe/Le Doc.

Avelino, Heriberto/Coler, Matt/Wetzels, Leo (edd.) (2015), *The Phonetics and Phonology of Laryngeal Features in Native American Languages*, Leiden, Brill.

Azeredo, José Carlos S. de (2008), *Gramática Houaiss da língua portuguesa*, São Paulo, Publifolha/IAH.

Bagno, Marcos (2011), *Gramática pedagógica do português brasileiro*, São Paulo, Parábola.

Barros, Maria Cândida D. M. (1993), *Os Intérpretes Jesuítas e a Gramática Tupi no Brasil (século XVI)*, Cadernos do Museu Paraense Emílio Goeldi 4.

Bessa Freire, José Ribamar (2004), *Rio Babel. A história das línguas na Amazônia*, Rio de Janeiro, EDUERJ/Atlântica.

Bonvini, Emilio (2008), *Línguas africanas e português falado no Brasil*, in: José Luiz Fiorin/Margarida Petter (edd.), *África no Brasil*, São Paulo, Contexto, 15–62.

Callou, Dinah Isensee (ed.) (1992), *A linguagem falada culta na cidade do Rio de Janeiro. Materiais para seu estudo*, vol. 1: *Elocuções Formais*, Rio de Janeiro, UFRJ/FJB.

Callou, Dinah M. Isensee/Duarte, Maria Eugênia Lamoglia (edd.) (2002), *Para a História do Português Brasileiro*, vol. 4: *Notícias de corpora e outros estudos*, Rio de Janeiro, Universidade Federal do Rio de Janeiro/Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro.

Callou, Dinah M. Isensee/Lobo, Tânia (edd.) (2020), *História do Português Brasileiro*, vol. 9: *História social do Português Brasileiro: da história social para a história linguística*, São Paulo, Contexto.

Callou, Dinah Maria Isensee/Lopes, Célia Regina (edd.) (1993), *A linguagem falada culta na cidade do Rio de Janeiro, materiais para seu estudo*, vol. 2: *Diálogo entre informante e documentador*, Rio de Janeiro, UFRJ/CAPES.

Callou, Dinah Maria Isensee/Lopes, Célia Regina (edd.) (1994), *A linguagem falada culta na cidade do Rio de Janeiro, materiais para seu estudo*, vol. 3: *Diálogos entre dois informantes*, Rio de Janeiro, UFRJ/CAPES.

Câmara Jr., Joaquim Mattoso (1939), *Lições de Linguística Geral I*, Revista de Cultura 25/146, 99–104.

Câmara Jr., Joaquim Mattoso ([4]1954), *Princípios de Linguística Geral como fundamento para os estudos superiores da língua portuguesa*, Rio de Janeiro, Acadêmica; ([1]1941), Rio de Janeiro, Briguiet.

Castilho, Ataliba T. de (ed.) (1970), *Projeto de estudo da norma linguística urbana culta no Brasil*, Marília, Conselho Municipal de Cultura.

Castilho, Ataliba T. de (1984), *El Proyecto de Estudio Coordinado de la Norma Culta. Formalismo y semanticismo en la sintaxis verbal*, in: Donald F. Solá (ed.), *Language in the Americas*, Proceedings of the Ninth PILEI Symposium, Ithaca, Cornell University, 161–165.

Castilho, Ataliba T. de (ed.) (1989), *Português Culto Falado no Brasil*, Campinas, Unicamp.

Castilho, Ataliba T. de (1990a), *O Português Culto Falado no Brasil. História do Projeto NURC/SP*, in: Dino Preti/Hudinílson Urbano (edd.), *A linguagem falada culta na cidade de São Paulo*, vol. 4: *Estudos*, São Paulo, Taq/Fapesp, 141–202.
Castilho, Ataliba T. de (ed.) (1990b), *Gramática do Português Falado*, vol 1: *A ordem*, Campinas, Unicamp/Fapesp.
Castilho, Ataliba T. de (ed.) (1993), *Gramática do Português Falado*, vol 3: *As abordagens*, Campinas, Unicamp/Fapesp.
Castilho, Ataliba T. de (1997a), *Língua falada e gramaticalização*, Filologia e Linguística Portuguesa 1, 107–120.
Castilho, Ataliba T. de (1997b), *Para uma sintaxe da repetição. Língua falada e gramaticalização*, Língua e Literatura 23, 293–332.
Castilho, Ataliba T. de (1998a), *Langue parlée et processus grammaticaux*, in: Mireille Bilger/Karel van den Eynde/François Gadet (edd.), *Analyse linguistique et approches de l'oral. Recueil d'études offert en hommage à Claire Blanche-Benveniste*, Paris/Leuven, Peeters, 141–148.
Castilho, Ataliba T. de (1998b), *A língua falada no ensino do português*, São Paulo, Contexto.
Castilho, Ataliba T. de (ed.) (1998c), *Para a História do Português Brasileiro*, vol. 1: *Primeiras ideias*, São Paulo, Humanitas/Fapesp.
Castilho, Ataliba T. de (2009), *An approach to language as a complex system. New issues in Historical Linguistics*, in: Vanderci Aguilera (ed.), *Para a História do Português Brasileiro*, vol. 7: *Vozes, veredas, voragens*, tomo 1, Londrina, Universidade Estadual de Londrina, 1–44.
Castilho, Ataliba T. de (2010), *Nova Gramática do Português Brasileiro*, São Paulo, Contexto.
Castilho, Ataliba T. de (ed.) (2018), *História do Português Brasileiro*, vol. 1: *O Português Brasileiro em seu contexto histórico*, São Paulo, Contexto/Fapesp.
Castilho, Ataliba T. de (ed.) (2019a), *História do Português Brasileiro*, vol. 2: *Corpus diacrônico do Português Brasileiro*, São Paulo, Contexto/Fapesp.
Castilho, Ataliba T. de (ed.) (2019b), *História do Português Brasileiro*, vol. 5: *Mudança sintática das construções: perspectiva funcionalista*, São Paulo, Contexto/Fapesp.
Castilho, Ataliba T. de (2019c), *Produção científica do Projeto para a História do Português Brasileiro*, unpublished.
Castilho, Ataliba T. de/Basílio, Margarida (edd.) (1996), *Gramática do Português Falado*, vol 4: *Estudos Descritivos*, Campinas, Universidade Estadual da Unicamp/Fundação de Amparo à Pesquisa do Estado de São Paulo.
Castilho, Ataliba T. de/Elias, Vanda Maria (2012), *Pequena Gramática do Português Brasileiro*, São Paulo, Contexto.
Castilho, Ataliba T. de, et al. (2019), *Diacronia da concordância*, in: Ataliba T. de Castilho (ed.), *História do Português Brasileiro*, vol. 5: *Mudança sintática das construções: perspectiva funcionalista*, São Paulo, Contexto/Fapesp, 284–400.
Castilho, Ataliba T. de/Preti, Dino (edd.) (1986), *A linguagem falada culta na cidade de São Paulo, materiais para seu estudo*, vol. 1: *Elocuções formais*, São Paulo, Taq/Fapesp.
Castilho, Ataliba T. de/Preti, Dino (edd.) (1987), A *linguagem falada culta na cidade de São Paulo, materiais para seu estudo*, vol. 2: *Diálogos entre dois informantes*, São Paulo, Taq/Fapesp.
Cavaliere, Ricardo (2012), *Gramaticografia da língua portuguesa no Brasil: tradição e inovação*, Limite. Revista de estudios portugueses y de lusofonía 6, 217–236.
Christino, Beatriz (2001), *Critérios de aceitabilidade gramatical e purismo na década de 1920*, Estudos Linguísticos 30, http://www.gel.org.br/estudos-linguisticos/edicoes.php, last accessed 13.02.2022.
Coelho, Adolpho (1881), *Os dialetos românicos ou neo-latinos na África, Ásia e América*, Boletim da Sociedade de Geografia de Lisboa.

Coelho, Olga (1998), *Serafim da Silva Neto (1917–1960) e a Filologia Brasileira. Um ensaio historiográfico sobre o papel da liderança na articulação de um paradigma em ciência da Linguagem*, Doctoral thesis, São Paulo, Universidade de São Paulo.

Coelho, Olga (2012), *O Português do Brasil em Macedo Soares (1938–1905)*, Límite. Revista de estudios portugueses y de lusofonía 6, 199–215.

Coelho, Olga/Danna, Stela Maria D. G./Polachini, Bruna Soares (2014), *O Português do Brasil em gramáticas brasileiras do século XIX*, Confluência 46, 115–140, http://www.gutenberg.org/files/33159/33159-h/33159-h.htm, last accessed 13.02.2022.

Coseriu, Eugenio (1976), *Perspectivas gerais*, transl. by Maria Candida Diaz Bordenave/Marilda Winkler, in: Anthony Naro (ed.), *Tendências atuais da linguística e da filologia no Brasil*, Rio de Janeiro, Francisco Alves, 11–44; ([1]1968), *General Perspectives*, in: Robert Lado/Norman A. McQuown/Sol Saporta (edd.), *Ibero-American and Caribbean Linguistics*, The Hague, Mouton, 5–62.

Couto de Magalhães, José Vieira (1876), *O Selvagem*, Rio de Janeiro, Typ. da Reforma.

Cyrino, Sonia M. L./Torres Morais, Maria Aparecida C.R. (edd.) (2018), *História do Português Brasileiro*, vol. 6: *Mudança sintática do Português Brasileiro: perspectiva gerativista*, São Paulo, Contexto/Fapesp.

Dias, Pedro (1697), *Arte da lingoa de Angola, oeferecida a Virgem Senhora N. Do Rosario, Mãy, e Senhora dos mesmos Pretos*, Lisboa, Miguel Deslandes.

Dias de Moraes, Lygia Corrêa (1987), *Nexos de Coordenação na Fala Urbana Culta de São Paulo*, Doctoral thesis, São Paulo, Universidade de São Paulo.

Dietrich, Wolf (1980), *Bibliografia da Língua Portuguesa do Brasil*, Tübingen, Narr.

Figueira, Luís (1621), *Arte da Lingua Brasílica*, Lisboa, Manoel da Silva.

Fiorin, José Luiz/Petter, Margarida (2008), *África no Brasil. A formação da língua portuguesa*, São Paulo, Contexto.

França, Angela Maria Ribeiro (1995), *Levantamento das proposições dos autores citados nos Princípios de Linguística Geral*, São Paulo, Universidade de São Paulo, CNPq/Departamento de Linguística, unpublished.

Franchi, Carlos (1977), *Linguagem, atividade constitutiva*, Cadernos de estudos linguísticos 22, 9–41. [Reprint in: Carlos Franchi/José Luiz Fiorin/Rodolfo Ilari (2011), *Linguagem, atividade constitutiva. Teoria e poesia*, São Paulo, Parábola].

Gândavo, Pero de Magalhães (1858, [1]1576), *História da Provincia Sãta Cruz, a que vulgarmête chamamos Brasil*, Revista do Instituto Historico e Geographico Brasileiro 21, 367–430.

Gildea, Spike/Queixalós, Francesc (2010), *Ergativity in Amazonia*, Amsterdam/Philadephia, Benjamins.

Hilgert, José Gaston (ed.) (1997), *A linguagem falada culta na cidade de Porto Alegre*, vol. 1: *Diálogos entre informante e documentador*, Passo Fundo/Porto Alegre, Ediupf/Ufrgs.

Hora, Dermeval da/Battisti, Elisa/Monaretto, Valéria de Oliveira (edd.) (2019), *História do Português Brasileiro*, vol. 3: *Mudança fonológica do Português Brasileiro*, São Paulo, Contexto.

Hora, Dermeval da/Rosa, Camilo (edd.) (2010), *Para a História do Português Brasileiro*, vol. 8, João Pessoa, Universidade Federal da Paraíba.

Ilari, Rodolfo (ed.) (1992), *Gramática do Português Falado*, vol 2: *Níveis de análise*, Campinas, Universidade Estadual de Campinas/Fundação de Amparo à Pesquisa do Estado de São Paulo.

Ilari, Rodolfo (ed.) (2014), *Gramática do português culto falado no Brasil*, vol. 3: Palavras de classe aberta, São Paulo, Contexto.

Ilari, Rodolfo (ed.) (2015), Gramática do português culto falado no Brasil, vol. 4: *Palavras de classe fechada,* São Paulo, Contexto.

Ilari, Rodolfo/Basso, Renato (2020), *História do Português Brasileiro*, vol. 11: *Semântica diacrônica do Português Brasileiro*, São Paulo, Contexto.
Jancsó, István, et al. (1994), *Cronologia de História do Brasil Colonial (1500–1831)*, São Paulo, Universidade de São Paulo, Departamento de História.
Jubran, Clélia C. S. (ed.) ([2]2015), *Gramática do Português Culto Falado no Brasil*, vol. 1: *A construção do texto falado*, São Paulo, Contexto.
Kato, Mary Aizawa (ed.) (1996), *Gramática do Português Falado*, vol 5: *Convergências,* Campinas, Universidade Estadual de Campinas/Fundação de Amparo à Pesquisa do Estado de São Paulo.
Kato, Mary A./Nascimento, Milton (edd.) (2015), *Gramática do Português Falado*, vol. 3: *A construção da sentença*, São Paulo, Contexto.
Kewitz, Verena (2007), *Gramaticalização e semanticização das preposições "a" e "para" no português brasileiro (séculos XIX e XX)*, Doctoral thesis, São Paulo, Universidade de São Paulo.
Koch, Ingedore Grunfeld Vilaça (ed.) (1996), *Gramática do Português Falado*, vol. 6: *Desenvolvimentos,* Campinas, Universidade Estadual de Campinas/Fundação de Amparo à Pesquisa do Estado de São Paulo.
Léry, Jean de (1980), *Viagem à Terra do Brasil*, São Paulo, EDUSP/Belo Horizonte, Itatiaia; ([1]1578), *Histoire d'un voyage fait en la Terre du Bresil*, La Rochelle, Antoine Chuppin.
Lima, Suzi/Rothstein, Susan (2020), *A typology of the mass/count distinction in Brazil and its relevance for mass/count theories*, Amsterdam/Philadephia, Benjamins.
Lobo, Tânia C. F., et al. (edd.) (2006), *Para a História do Português Brasileiro*, vol. 6: *Novos dados, novas análises*, 2 tomos, Salvador, Universidade Federal da Bahia.
Lopes, Célia R. dos S. (ed.) (2018), *História do Português Brasileiro*, vol. 4: *Mudança sintática das classes de palavras: perspectiva funcionalista*, São Paulo, Contexto/CNPq.
Macedo Soares, Antonio Joaquim de (1880), *Algumas palavras africanas introduzidas no português que se fala no Brasil*, in: Pinto (ed.) (1978), 45–47.
Macedo Soares, Antonio Joaquim de (1889), *Diccionario Brazileiro da Língua Portugueza, elucidario etymologico-critico das palavras e phrases que, originarias do Brazil, ou aqui populares, se não encontrão nos diccionarios da lingua portugueza, ou nelles vêm com forma ou significação differente 1875–1888*, Rio de Janeiro, G. Leuzinger & Filhos, 300–444.
Machado, José Pedro (1956), *Dicionário Etimológico da Língua Portuguesa*, vol. 1, Lisboa, Editorial Confluência
Machado, José Pedro (1977), *Dicionário Etimológico da Língua Portuguesa*, vol. 2, Lisboa, Livros Horizonte.
Magalhães, Erasmo d'Almeida (1974), *Quinze anos de Linguística Indígena Brasileira*, Língua e Literatura 3, 251–278.
Magalhães, Erasmo d'Almeida (1975), *Bibliografia de Linguística Indígena Brasileira (1954–1974)*, Língua e Literatura 4, 149–184.
Magalhães, Erasmo d'Almeida (1981), *As atividades do Summer Institute of Linguistics no Brasil*, Biblos 57, 753–772.
Mamiani, Luis Vincencio (1699), *Arte de grammatica da lingua brasilica da naçam Kiriri*, Lisboa, Miguel Deslandes.
Marcuschi, Luiz Antonio (1983), *Análise da conversação*, São Paulo, Ática.
Marcuschi, Luiz Antonio (1991), *Análise da conversação e análise gramatical*, Boletim da Abralin 10, 11–34.
Marcuschi, Luiz Antonio (1997), *Fala e escrita no continuum tipológico*, Recife, Universidade Federal de Pernambuco.
Martius, Karl F. P. von (1969, [1]1867), *Wörtersammlung Brasiliannischer sprachen/Glossaria Linguarum Brasiliensium/Glossários de Diversas Línguas e Dialectos, que Fallao os Indios no Império do Brazil*, Wiesbaden, Martin Sändig.

Mattos e Silva, Rosa Virgínia (ed.) (2001), *Para a História do Português Brasileiro*, vol. 2: *Primeiros Estudos*, 2 tomos, São Paulo, Humanitas/Fapesp.

Maurer Jr., Theodoro Henrique (1952), *Introdução* [to the study of Hawking W. Neil. *A Fonologia da Língua Uáiuái*.], Boletim da Faculdade de Filosofia, Ciências e Letras da Universidade de São Paulo 1/57, 1–5.

Módolo, Marcelo (2004), *Gramaticalização e semanticização das orações correlativas no português*, Doctoral thesis, São Paulo, Universidade de São Paulo.

Moraes de Castilho, Célia Maria (2001), *Seria quatrocentista o Português implantado no Brasil? Estruturas sintáticas duplicadas em textos portugueses do séc. XV*, in: Rosa Virginia Mattos e Silva (ed.) (2001), *Para a História do Português Brasileiro*, vol. 2: Primeiros Estudos, tomo 1, 57–89.

Moraes de Castilho, Céla Maria (2013). *Fundamentos sintáticos do português brasileiro*, São Paulo, Contexto.

Motta, Jacyra/Rollemberg, Vera (edd.) (1994), *A linguagem falada culta na cidade de Salvador, materiais para seu estudo*, vol. 1: *Diálogos entre informante e documentador*, Salvador, Instituto de Letras da UFBa.

Motta, Jacyra/Rollemberg, Vera (edd.) (2006), *A linguagem falada culta na cidade de Salvador, materiais para seu estudo*, vol. 2: *Elocuções formais*, Salvador, Instituto de Letras da UFBa/ CNPq.

Naro, Anthony Julius/Scherre, Maria Marta P. (edd.) (2007), *Origens do português brasileiro*, São Paulo, Parábola.

Nascentes, Antenor (1922), *O linguajar carioca*, Rio de Janeiro, Organização Simões.

Nascimento, Mílton do (1993), *Gramática do Português Falado: articulação teórica*, in: Ana Maria Stahl Zilles (ed.) (2005), *Estudos de variação linguística no Brasil e no ConeSul*, Porto Alegre, Universidade Federal do Rio Grande do Sul, 93–116.

Neves, Maria Helena de Moura (ed.) (1999), *Gramática do Português Culto Falado no Brasil*, vol. 7, Campinas, São Paulo, Unicamp/Fapesp/Humanitas.

Neves, Maria Helena de Moura (2000), *Gramática de usos do português*, São Paulo, Unesp.

Neves, Maria Helena de Moura (ed.) (2016), *Gramática do português culto falado no Brasil*, vol. 5: *A construção das sentenças hipotáticas*, São Paulo, Contexto.

Neves, Maria Helena de Moura (2018), *A gramática do português revelada em textos*, São Paulo, Unesp.

Oliveira Jr., Miguel (2019), *NURC 50 anos: 1969–2019*, São Paulo, Parábola.

Paiva, Vera Lúcia M. de Oliveira/Nascimento, Milton do (edd.) (2009), *Sistemas adaptativos complexos. Lingua(gem) e aprendizagem*, Belo Horizonte, Faculdade de Letras da UFMG.

Penhavel, Eduardo/Cintra, Marcos Rogério (edd.) (in print), *História do Português Brasileiro*, vol. 8: *Abordagem diacrônica dos processos de construção textual do Português Brasileiro*, São Paulo, Contexto.

Perini, Mário Alberto (1995), *Gramática descritiva do português*, São Paulo, Ática.

Perini, Mário Alberto (2002), *Modern Portuguese: a reference gramar*, New Haven, Connecticut, Yale University Press.

Perini, Mário Alberto (2010), *Gramática do português brasileiro*, São Paulo, Parábola.

Perini, Mário Alberto (2016), *Gramática descritiva do português brasileiro*, Petrópolis, Vozes.

Pinto, Edith Pimentel (ed.) (1978), *O Português do Brasil: textos críticos e teóricos*, vol. 1: *1820/1920, fontes para a teoria e a história*, Rio de Janeiro/São Paulo, Livros Técnicos e Científicos/ Universidade de São Paulo.

Pinto, Edith Pimentel (ed.) (1981), *O Português do Brasil: textos críticos e teóricos, vol. 2: 1920/1945, fontes para a teoria e a história*, Rio de Janeiro/São Paulo, Livros Técnicos e Científicos/ Universidade de São Paulo.

Praxedes, Silvana Gurgel (2008), *O período dos estudos brasileiros dito científico na questão da colocação pronominal (1880–1920)*, Master's thesis, São Paulo, Universidade de São Paulo.
Preti, Dino (ed.) ([2]1995, [1]1993), *Análise de textos orais*, São Paulo, FFLCH/USP.
Preti, Dino (ed.) (1997), *O discurso oral culto*, São Paulo, Humanitas/FFLCH/USP.
Preti, Dino (ed.) (1998), *Estudos de língua falada. Variações e confrontos*, São Paulo, Humanitas.
Preti, Dino (ed.) (2000), *Fala e escrita em questão*, São Paulo, Humanitas.
Preti, Dino (ed.) (2002), *Interação na fala e na escrita*, São Paulo, Humanitas.
Preti, Dino (ed.) (2005), *Diálogos na fala e na escrita*, São Paulo, Humanitas.
Preti, Dino (ed.) (2006), *Oralidade em diferentes recursos*, São Paulo, Associação Editorial Humanitas.
Preti, Dino/Urbano, Hudinílson (edd.) (1989), *A linguagem falada culta na cidade de São Paulo, materiais para seu estudo*, vol. 3: *Diálogos entre o informante e o documentador*, São Paulo, Taq/Fapesp.
Preti, Dino/Urbano, Hudinílson (edd.) (1990), *A linguagem falada culta na cidade de São Paulo*, vol. 4: *Estudos*, São Paulo, Taq/Fapesp.
Ramos, Jânia/Alckmin, Mônica A. (edd.) (2007), *Para a História do Português Brasileiro*, vol. 5: *Estudos sobre mudança linguística e história social*, Belo Horizonte, Faculdade de Letras da Universidade Federal de Minas Gerais.
Ramos, Jânia/Oliveira, Marilza (edd.) (2021), *História do Português Brasileiro*, vol. 10: *Dialetação e povoamento: da história linguística à história social*, São Paulo, Contexto/Fapesp.
Ribeiro, Julio ([2]1913, [1]1881), *Grammatica Portugueza*, Rio de Janeiro, Francisco Alves.
Roberts, Ian/Kato, Mary (ed.) (1993), *Português brasileiro – uma viagem diacrônica*, Campinas, Universidade Estadual de Campinas.
Robins, Robert Henry (1967), *A Short History of Linguistics*, London/Bloomington, Longman/Indiana Univ. Press.
Rodrigues, Aryon (1994), *Línguas Brasileiras*. Para o conhecimento das línguas indígenas, São Paulo, Loyola.
Rodrigues, Aryon (1996), *As Línguas Gerais Sul-Americanas*, Papia 4/2, 6–18.
Rodrigues, Aryon (1997), *Descripción del Tupinambá en el período colonial: el arte de José de Anchieta*, in: Klaus Zimmermann (ed.), *La descripción de las lenguas amerindias en la época colonial*, Frankfurt am Main, Vervuert, Madrid, Iberoamericana, 371–400.
Rodrigues, Aryon (2000), *Hipótese sobre as Migrações dos três Subconjuntos Meridionais da Família Tupi-Guaraní*, in: *Proceedings of the II Abralin National Congress*, CD-ROM, Florianópolis, Abralin.
Rosa, Maria Carlota (2003), *A Língua Mais Geral do Brasil nos Séculos XVI e XVII*, in: Bessa Freire/Carlota Rosa (edd.), *Línguas Gerais. Política linguística e catequese na América do Sul no período colonial*, Rio de Janeiro, EDUERJ, 133–146.
Sá, Maria P. M., et al. (edd.) (1996), *A linguagem falada culta na cidade do Recife*, vol. 1: *Diálogos entre informante e documentador*, Recife, Universidade Federal de Pernambuco, Programa de Pós-Graduação em Letras e Linguística.
Sá, Maria P. M., et al. (edd.) (2005), *A linguagem falada culta na cidade do Recife*, vol. 2: *Elocuções formais*, Recife, Universidade Federal de Pernambuco, Programa de Pós-Graduação em Letras e Linguística.
Said Ali, Manuel (1895), *A colocação dos pronomes na linguagem corrente*, Revista Brasileira 1.
Simões, José da Silva (2007), *Sintaticização, discursivização, e semanticização das orações de gerúndio no português brasileiro*, Doctoral thesis, São Paulo, Universidade de São Paulo.
Souza, Gabriel Soares de ([3]1938, [1]1587), *Tratado descriptivo do Brasil em 1587*, São Paulo, Nacional/EDUSP.
Vitral, Lorenzo (2017), *Gramática inteligente do português do Brasil*, São Paulo, Contexto.

Tânia Lobo

2 The social history of Brazilian Portuguese

Abstract: This chapter deals with the external history of Brazilian Portuguese with a focus on social factors, establishing the fundamental parameters within which the internal history of the language must be studied. It outlines the changing sociolinguistic situations throughout the presence of the Portuguese language on the American continent. Appropriately periodized, the discussion covers where the language was spoken and acquired by whom and which functions it served in each of the periods and communicative situations, providing also demographic information whenever available. By these means, it singles out the most important aspects of sociolinguistic developments in Brazil, such as the overall multiethnic context, the formation of a nation, the different stages of teaching of Portuguese within the different educational systems, the general high mobility of groups of speakers, and the drastic increase of speakers during the 20th century.

Keywords: social history of Brazilian Portuguese, language contact, linguistic demography, literacy demography, sociolinguistic continuum, Jesuits, Lingua geral

1 Introduction

This chapter is dedicated to the linguistic social history of Brazil, focusing on Brazilian Portuguese, with the objective of outlining a system of fundamental coordinates that may be connected to the so-called internal history of the language. In outlining this system, it will cover global and profound changes which mainly took place on two levels over a long period. The first level refers to the "connection between the facts of land occupation, facts of demographic and linguistic distribution, and facts about the prevalence and disappearance of languages" (Houaiss 1985, 31–32). The justification for this approach lies in the fact that BP is a language which emerged amid a multilingual context, and linguistic contact was therefore one of the aspects comprising its formation. The second level concerns the issue of the social distribution of writing in Brazil. Along with the history of contact, it is a key question for analyzing the varieties of BP that comprise the complex sociolinguistic web of contemporary Brazil, which is defined by a continuum featuring, at one extreme, the more socially esteemed linguistic norms (urbane, supposedly more unitarian, supposedly direct descendants of European Portuguese, which were also molded by the influx of the normative patterns of writing that are disseminated through scholastic study), and at the other end, the more socially stigmatized linguistic norms (rural, supposedly more diversified and supposedly direct descendants of the Portuguese spoken as a second language by natives and black people who, along with their descendants, comprised the majority of

https://doi.org/10.1515/9783110405958-003
Despite careful production of this book, a wrong map was printed in the original version of this chapter (p. 55). This has now been corrected. We apologize for the mistake.

Brazil's population from the 16th century through the mid-1800s and were mostly outside the margins of the formal education system through the mid-1900s).

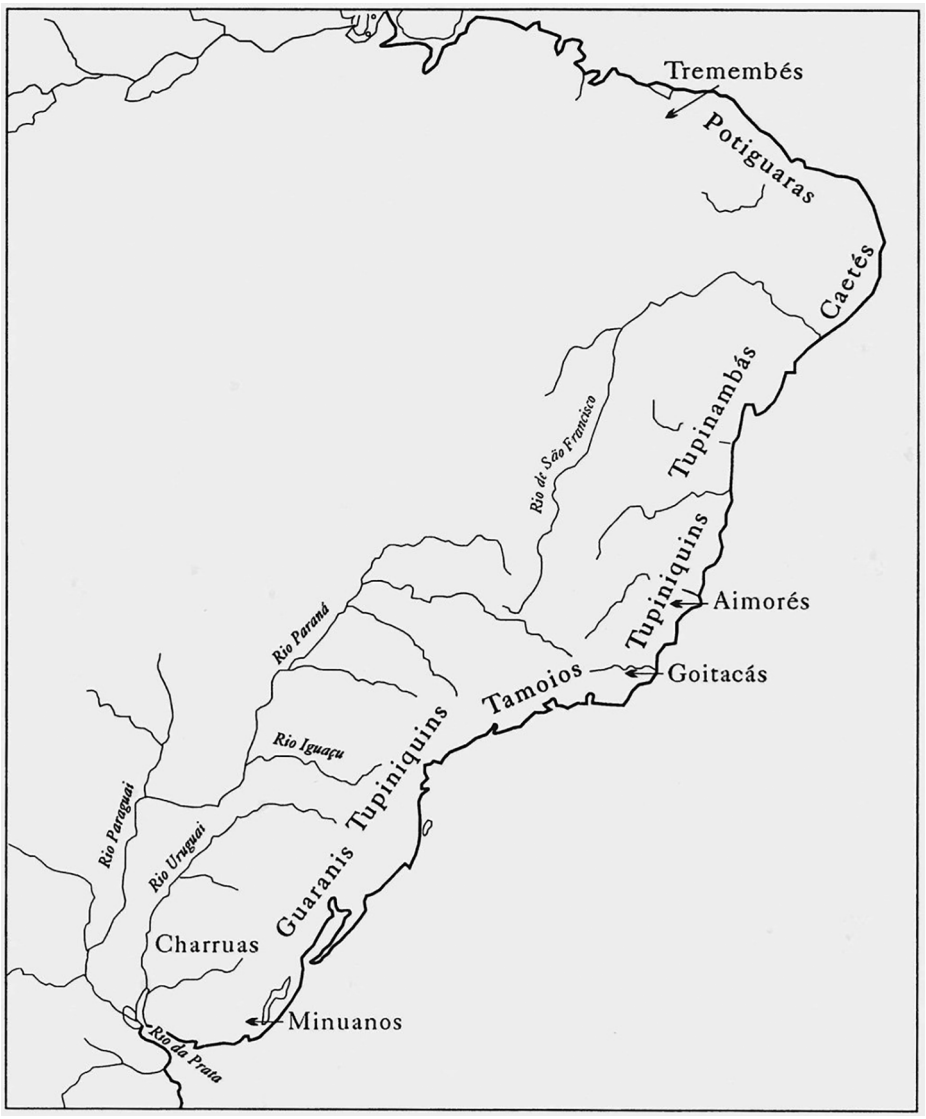
2 Origins of the diffusion of the Portuguese language throughout Brazil

2.1 The 16th century

The expansion of the Portuguese language outside of Europe began in the 15th century, and was inseparable from the adventures of the great voyages. In March of 1500, a fleet of ships, commanded by the nobleman Pedro Álvares Cabral, sailed out of Lisbon bound for India. They soon left the African coast, however, and in April they sighted the land that would be known as Vera Cruz, Santa Cruz, and finally Brazil.

There is consensus among scholars of Brazilian paleo-demography about the difficulty of calculating the indigenous population at the time that they began contact with Europeans. This difficulty is explained by the fragmentary nature of the archaeological remains and also by the few remaining records made by employees of the Portuguese crown or by the church in the latter decades of the 16th century. It is equally difficult to calculate the number of languages spoken by the natives at that time. The currently accepted estimate is that there were some 1,175 languages in Brazil before colonization began (Rodrigues 1993).

The coastal indigenous populations who first came into contact with the Portuguese spoke mostly languages in the Tupi-Guarani family, from the Tupi branch, which is the largest group of Brazilian indigenous languages, followed by the Macro-Ge trunk and the Arawak, Carib and Páno families. The map below shows how the Tupi-Guarani "nations" were distributed along the coast; it also indicates the points which were marked by the presence of ethnically and linguistically distinct groups, known as tapuias:



Map 1: Distribution of the Tupi-Guarani "nations" along the Brazilian coast in the early 16th century (source: Fausto 1992).

The passage below, extracted from the *Tratado da Província do Brasil* ('Treaty of the Province of Brazil') by Pero de Magalhães de Gândavo, who was in Brazil between 1558 and 1572, not only refers to the catastrophic decline of the coastal native peoples—who, because of wars and the spread of new diseases, were reduced to one third of their population by the end of the 1500s (Marcílio 1999, 315; Andreazza/Nadalin 2011, 60)—, but also to the linguistic homogeneity which had characterized them:

> 'There were many of these coastal Indigenous in the colonies; everywhere was full of them when the Portuguese began to settle the land: but, because these same Indigenous rose up against them and betrayed them many times, the governors and captains of the land destroyed them little by little and killed many of them. Others fled to the *sertão* [outback] and that was how the coast became depopulated by the heathens throughout the period of the *capitanias* (early colonies). Adjacent to these, there remained some villages of these Indigenous who are peaceful and friends of the Portuguese. All these heathens on the coast speak one language, one lacking three letters—scilicet—there is no f, l, or r, which is amazing, because there is also no faith, no law, and no ruler, and thus they live without justice and in a disorderly manner'[1] (Gândavo 1965, 179, our translation).

This language widely spoken on the coast—called Tupinambá, but commonly referred to as Tupi in the past—is one of the main languages of the Brazilian colonial history and was described by the Jesuit José de Anchieta in *Arte de gramática da língua mais usada na costa do Brasil* (1980) ('The art of grammar of the most commonly used language along the Brazilian coast'), the first grammar of a Brazilian indigenous language. It was printed in Coimbra in 1595. This grammar affirms the recurring practice among the Jesuits of religious indoctrination in the languages of the subjugated peoples. As for this 'most commonly used language on the coast of Brazil', it is relevant to inquire to what extent this linguistic homogeneity on the coast resulted from the genocide that devastated the coastal natives. In addition, since this language is often referred to as "the general language of the coast", it is also worth noting that the institutionalization of indigenous languages as "general" (in the sense of inter-ethnic vehicular languages) would have been part of a colonial indigenous policy that, disregarding any differences between groups, including linguistic differences, instituted the supra-ethnic leveling category *índio* 'Indian' (Barros/Borges/Meira 1996, 195–196).

The period between 1500 and 1530 was marked by the ephemeral establishment of factories and barter for brazilwood. It was only in the 1530s, during the reign of Dom

1 "Havia muitos destes índios pela costa junto das capitanias; tudo enfim estava cheio deles quando começaram os portugueses a povoar a terra: mas, porque os mesmos índios se levantavam contra eles e faziam-lhes muitas traições, os governadores e capitães da terra destruíram-nos pouco a pouco e mataram muitos deles. Outros fugiram para o sertão e assim ficou a costa despovoada de gentio ao longo das capitanias. Junto delas ficaram algumas aldeias destes índios que são de paz e amigos dos portugueses. A língua deste gentio toda pela costa é uma, carece de três letras—*scilicet*—não se acha nela f, nem l, nem r, coisa digna de espanto, porque assim não tem fé, nem lei, nem rei, e desta maneira vivem sem justiça e desordenadamente."

João III, that Portugal, which at that time had a population of between 1,000,000 and 1,500,000 inhabitants (Andreazza/Nadalin 2011, 57–58), decided to begin colonizing its share of the Americas. The first initiative was to send the exploratory expedition led by Martim Afonso de Souza (1530–1533), who established a center of Portuguese settlement in São Vicente in what is now the state of São Paulo. Next, the territory was split from North to South, from the coast up to the meridian of Tordesillas, into 15 captaincies. In addition, the sugar industry, which had already been tested and proven successful in the São Tomé and Príncipe archipelagos, and Madeira and the Azores, was extended to Brazil, focusing on the Northeast.

LINHA ÆQVINOCIAL.
Rio das Amazonas
Capitania de João de Barros
Capitania de Francisco barreto
Capitania de Jorge dalbuquerque
DE SVA MAGESTADE.
Capitania de Francisco giraldes
Capitania do Duque daveiro.
Capitania de Vasco fz Coutinho
Capitania de Pero de goes
Capitania de Joam de Sousa
LINHA DA DEMARCACAM
TROPICO DE CAPRICORNIO
Rio da prata
Estreito de Fernando de Magalhães
O C E A N O
M E R I D I O N A L

Map 2: Hereditary captaincies (Cintra 2013, 39).

One of the powers of the crown and of the holders of hereditary captaincies was the distribution of sesmarias—plots of agricultural land with dimensions varying from subsistence farms to large agro-exporting estates. So, although it has been stated that the plantation (a large agricultural property based on monoculture for export and sustained by slave labor) was the basic form of colonization of Brazil, the Portuguese motherland always sought to ensure conditions for food production within the colony itself. This statement points to the fact, which is relevant from the perspective of a linguistic history, that even in rural areas, social structure was not exclusively reduced to the polarization of the lords *versus* the enslaved, since there was a significant portion of the population consisting of poor white peasants (Fausto 1995, 58–59).

At the beginning of colonization, the movement of migrants from Europe was characterized as spontaneous yet selective: adult men of different social strata predominated, coming mostly from the Northwest of Portugal and the Atlantic islands, as well as a considerable number of new Christians (Jews required to "convert" to Christianity). The almost exclusive presence of men, from the beginning, fostered intense racial mixing, which was fundamental to the creation of the Brazilian population. It is also relevant from the perspective of a linguistic history, since mixed-race peoples are a link between linguistically-distinct worlds.

The 1530s saw another predominantly male migratory movement, which was also selective but not spontaneous: the arrival of enslaved Black Africans. Brazil was the last country to abolish slavery in the West, only in 1888. Slave trade remained legal until 1850, and although it is not possible to precisely determine how many were brought to Brazil, it is estimated that between 8,000,000 and 11,000,000 enslaved Africans were brought to the Americas, and approximately 4,900,000 of this total were brought to Brazil (Schwarcz/Starling 2015, 82). In the 16th century, Portuguese smugglers used the term "Guinea negro" to describe Africans. In an inquiry into what Guinea was in the early days of slave trade, Oliveira (1997) reported that all of West Africa north of the Equator, from the Senegal River to Gabon, was then known as Guinea, and stated that the term also came to be applied to populations south of the Equator. It can be concluded, therefore, that since the beginning, languages were transplanted from two sub-Saharan regions, and would characterize the entire history of slave trade in Brazil: those of West Africa, including languages ranging from Senegal to Nigeria, and Bantu, with languages spanning the entire continent south of the Equator. The languages of the African continent are classified into four groups: Congo-Kordofanian, Nilo-Saharan, Afroasiatic, and Khoisan; Congo-Kordofanian is subdivided, in turn, into two families: Niger-Congo and Kordofanian.[2] The enslaved people that came to Brazil spoke languages of the Niger-Congo family, predominantly of the Bantu and Kwa groups (Castro 2001).

2 cf. Greenberg (1966) apud Castro (2001).

Map 3: Regions where slave trade was concentrated (source: Castro 2001, 46).

Motivated by internal factors—the precarious administration of several captaincies—as well as external ones—the beginning of the trade crisis with the East, defeats suffered in Africa, and the Spanish discovery of a huge silver mine in Potosí, Peru in 1545 (Fausto 1995, 46)—, Dom João III decided to install a general government in Brazil in 1549, and sent the nobleman Tomé de Sousa to be the first governor. He founded Salvador, the first colonial capital. His entourage included five Jesuits, led by the

priest Manuel da Nóbrega. From 1549 to 1759, when they were driven out of the vast Portuguese empire, the Jesuits held a monopoly on education in Brazil.[3] It fell to Nóbrega to formulate the colony's first curriculum, which was directed not only at the children of the Portuguese, but also at the indigenous children:

> 'The plan began with learning Portuguese (for the native people) and continued with Christian doctrine, schooling to teach reading and writing, and, optionally, choral singing and instrumental music. It culminated on one side in professional and agricultural learning, and on the other side with Latin grammar for those who were destined for higher studies in Europe (at the University of Coimbra)'[4] (Saviani 2013b, 43, our translation).

Despite the careful deployment of this syllabus through a network of high schools and despite having initiated a strategy to bring orphan boys from Lisbon as a way to "attract indigenous boys and, through them, to act upon their parents, especially the chiefs, converting the whole tribe to the Catholic faith", Nóbrega had little success. The plan's opponents included the Jesuits themselves (Saviani 2013b, 41), and the syllabus was replaced in 1599 by the *Ratio atque Institutio Studiorum Societas Jesu*:

> 'The plan contained in the *Ratio* was universalist and elitist in nature. It was universalist because the plan was considered universally accepted by all the Jesuits, no matter where they were; it was elitist because it was meant only for the children of the settlers, not the Indigenous, with the Jesuit schools becoming the tool for forming the colonial elite. In doing so, the initial stages established in the Nóbrega plan (learning Portuguese and how to read and write) were eliminated.'[5] (Saviani 2013b, 56, our translation).

A thought-provoking question relates to what extent writing was widespread in the diminutive colonial society of the late 16th century, when the Ratio Studiorum was being implemented. Because of the lack of official figures, attempts to approximate literacy numbers in Ancien Régime societies have been carried out by calculating signatures. Analyzing the statements made and signed before the Inquisition on its visit to Brazil, specifically to the captaincies of Bahia, Pernambuco, Itamaracá, and Paraíba between

---

3 Other religious orders carried out educational activities in colonial Brazil, but they 'operated in a scattered and intermittent manner, with no official support or protection [...]' (Saviani 2013b, 41).

4 "O plano iniciava-se com o aprendizado do português (para os indígenas); prosseguia com a doutrina cristã, a escola de ler e escrever e, opcionalmente, canto orfeônico e música instrumental; e culminava, de um lado, com o aprendizado profissional e agrícola e, de outro lado, com a gramática latina para aqueles que se destinavam à realização de estudos superiores na Europa (Universidade de Coimbra)".

5 "O Plano contido no *Ratio* era de caráter universalista e elitista. Universalista porque se tratava de um plano adotado indistintamente por todos os jesuítas, qualquer que fosse o lugar onde estivessem. Elitista porque acabou destinando-se aos filhos dos colonos e excluindo os indígenas, com o que os colégios jesuítas se converteram no instrumento de formação da elite colonial. Por isso, os estágios iniciais previstos no Plano de Nóbrega (aprendizado de português e escola de ler e escrever) foram suprimidos."

1591 and 1595, Lobo/Oliveira (2013) and Sartori (2016) reached the following general conclusions: in a universe of 686 subjects, 492 (71 %) signed their testimonies and 194 (28 %) did not. The surprising rate of 71 % of signatures can be better interpreted by redistributing the sample by variables, with sex and ethnicity being two of the most important:

Table 1: Signature rates compared with sex and ethnicity (source: Lobo and Oliveira 2013; Sartori 2016).

		Subscribers	Not subscribers	Total
Whites	Men	424 – 92,8 %	33 – 7,2 %	457 – 100 %
	Women	18 – 13,7 %	113 – 86,3 %	131 – 100 %
Indios	Men	–	2 – 100 %	2 – 100 %
	Women	–	2 – 100 %	2 – 100 %
Black people	Men	–	1 – 100 %	1 – 100 %
	Women	–	3 – 100 %	3 – 100 %
Mamelucos	Men	32 – 100 %	–	32 – 100 %
	Women	4 – 17,4 %	19 – 82.6 %	23 – 100 %
Mulattos	Men	4 – 80 %	1 – 20 %	5 – 100 %
	Women	–	6 – 100 %	6 – 100 %
Cafuzos	Men	2 – 67 %	1 – 33 %	3 – 100 %
	Women	–	2 – 100 %	2 – 100 %
Gypsies	Men	–	1 – 100 %	1 – 100 %
	Women	–	7 – 100 %	7 – 100 %
Not identified	Men	8 – 100 %	–	8 – 100 %
	Women	–	3 – 100 %	3 – 100 %
Total	Men	470 – 93,3 %	39 – 7,7 %	509 – 100 %
	Women	22 – 12,4 %	155 – 87,6 %	177 – 100 %

Although these evidence-based data are somewhat limited, they do allow for some observations: 1) as expected, the signature rates were almost inverted between men and women: while 93.3 % of the men signed, 87.6 % of women did not; 2) the male sample was basically comprised of whites (89.7 %) and *mamelucos* (mixed-race children of white men and indigenous women)—(6.3 %); among white men, 92.8 % were able to sign, and among the *mamelucos*, 100 %; 3) despite the known limits of the signature calculation method and the need to discuss the representativeness of the sample, the data allow for the hypothesis that, during initial colonization, especially in Bahia and

Pernambuco, white Portuguese men and their white and mixed-race descendants comprised a select group of individuals who were familiar with the technology of writing, which was indispensable for communication with the distant Portuguese motherland. And this is what forms the foundation of linguistic "duality", which would become one of the distinctions of Brazilian colonial and imperial society.

At the end of the 16th century, just more than half of the 15 original captaincies were functioning. Although the figures presented are "crude and merely indicative estimates" (Couto 1998, 277), the table below facilitates an approximation of the 16th-century population profile, as well as some linguistic inferences:

Table 2: Evolution of the Brazilian population in the 16th century, by captaincy (source: Couto 1998, 276–277).

	1546			1570			1585			1590		
	PORT	**IND**	**AFR**	**PORT**	**IND**	**AFR**	**PORT**	**IND**	**AFR**	**PORT**	**IND**	**AFR**
Paraíba[6]	–	–	–	–	–	–	–	–	–	825	–	400
Itamaracá	–	–	–	550	–	–	275			495	–	250
Pernambuco	3025	500		5500	–	–	8000	2000	10000	11.000	2000	18.000
Bahia	1100	260		6050	–	–	11000	8000	3000	8250	3600	18.000
Ilhéus	330	80		1100	–	–	825	–	–	1650	2000	400
P. Seguro	–	–	–	1210	–	–	550	–	–	1595	3000	3000
E. Santo	1650	300		1100	–	–	825	4500	–	2200	9000	700
Rio de Janeiro	–	–	–	770	–	–	825	3000		1540	3000	700
S. Vicente/ S. Amaro	3300	500	–	2750	–	–	1650	–	–	3300	6000	800
TOTAL	9405	1640	–	19030	–	–	23950	17500	13000	30855	28600	42250

In the year 1590, which has less fragmented data, the Brazilian population was 101,705 inhabitants, distributed unevenly among nine captaincies. Pernambuco and Bahia were equal and had strong demographic concentration, and together were home to 60 % of the colonial population. Of this total, 30.3 % were Portuguese, 28.1 % were Natives, and 41.5 % were Africans. The first aspect to highlight here is the contrast between approximately 30 % being "whites" and 70 % "non-whites", which was one of the constants in Brazilian demographic history until the mid-19th century. This shows that understanding the process of diffusion of the Portuguese language in

6 The French conquered Paraíba in 1585, which is why this colony only appears in the table in 1590.

Brazil cannot disregard the role played by Natives and black people, who acquired the language of the colonizer as a second language, to differing degrees of proficiency, and passed it on to their descendants.

Africans, who were present in all the captaincies, made up the majority of the population with 41.5%; in Pernambuco and Bahia, they accounted for 58.1% and 64.2%, respectively. Two fundamental questions to be addressed later are why no African languages have continued to be spoken until the present day and why, despite the presence of what were certainly endemic processes of pidginization, no creole language was established in Brazil.

Finally, the table brings attention to the fact that racial mixing is not represented, as during the period there were recurring requests for the mother country to send orphaned women and banished men, who were not thieves, to stop or minimize racial mixing. As for the *mamelucos*, according to Couto (1998, 275), in the majority of sources, these were included among Portuguese families[7] and consequently comprised a group of "non-Europeans" that cannot be counted. Thus, this represents what has been called the "whitening" of mixed-race people; this group of *mamelucos* was "whitened" in the 16th century in Bahia and Pernambuco, not only by speaking the language of their European parents, but by writing it.

2.2 The 17th century

While the 16th century was characterized by the occupation of central areas of the coast, the 17th was marked by progressive advancement inland (to the *sertão* badlands) and by occupation of peripheral swaths of the coast. In the North, this included parts of Maranhão and Pará, from where the colonization of Amazonia began, and in the South, areas of what are now the states of Paraná and Santa Catarina. Two separate colonies were defined in South America, which had little contact with each other and also different linguistic histories: the State of Brazil, which covered areas of the Northeast, Southeast, and South, and the State of Grão-Pará and Maranhão, which effectively was separated from the State of Brazil in 1624, and corresponds to the modern states of Piauí, Maranhão, Pará, Amazonas, Rondônia, Roraima e Amapá. It was only in the year 1823, after the independence of (the State of) Brazil in 1822, that the State of Grão-Pará and Maranhão was incorporated into imperial Brazil.

The migratory movement from Europe toward Brazil continued to mostly come from Portugal and to be spontaneous and selective, although there were three situations that departed from this rule (Marcílio 1999, 320–321): the organized arrival of families from the Azores and Madeira who settled on the coast of Maranhão and Pará to protect it against raids; greater freedom of entry for Catholic Europeans of any nation-

7 The inquisitorial sources are one of the exceptions, as seen above.

ality during the period of Spanish domination (1580–1640), when Spanish immigrants settled in peripheral areas of the South; and finally, as a result of invasions, the Dutch presence in the major areas of sugar production—Bahia and Pernambuco—in the years 1624–1625 and 1630–1654. None of these three events had relevant linguistic consequences, clearly because the three groups had no significant demographic impact.

As for the black people, in the 1600s, slave trade was concentrated almost exclusively in the Congo and in Angola. The overwhelming predominance of Bantu language speakers then resulted in the writing of a Kimbundu grammar in Bahia by the Jesuit Pedro Dias, *A arte da língua de Angola (The art of the language of Angola)*, which was published in Lisbon in 1697. Jesuit Antonio Vieira affirmed the extent of the use of Kimbundu (which he designated as "Ethiopian") by saying it was the "language in which, within this city [Salvador] alone, they taught and catechized twenty-five thousand black people, not to mention countless numbers outside [...]" (Silva Neto 1986, 75). However, the 25,000 or more black people mentioned by Vieira would not have spoken only Kimbundu; since they came from various regions of Angola, the role of the Kimbundu language as a vehicle language can be inferred (Castro 2001, 55). The three languages in the Bantu group with most significant representation in Brazil were Kimbundu, Kongo, and Umbundu, and it is possible that there was intercommunication among their speakers. From the 17th century, the speakers of Bantu languages were always in the majority: from 1580 to 1690, they corresponded to 93 % of the enslaved population; from 1691 to 1750, 55 %; from 1751 to 1808, 68 %, and finally from 1808 to 1850, 71 % (Almeida 2014, 353–355). It is consequently worth noting, with respect to contact between Portuguese and African languages, that because of their long history in Brazil and the population density and geographic range attained by their speakers, the Bantu languages generally had a much more profound linguistic influence than the other ethno-linguistic groups (Castro, undated, 2).

atividade principal	século de introdução maciça			
	XVI	XVII	XVIII	XIX
agricultura	B	B/J	B/J/N	B/J/N
mineração			B/J	
serviços urbanos				B/J/N/H

Map 4: African ethnological map outline in Brazil (source: Castro 2001, 47).

Moving past the impact of the initial contact phase, the history of the indigenous peoples (and, consequently, their languages) came to be identified by complex and drastic terms: extermination, which was not confined to the 16th century and extends dramatically to the present day; subjugation, either as enslaved people working on farms

or plantations[8] or as "free" people in allied villages or in Jesuit settlements; so-called "integration" (through broad racial mixing); and even multiple forms of resistance, whether this meant fleeing to the interior, or as renewed historiography of the indigenous Americas has shown, by constant and complex processes of redefining identity.

The territorial expansion was motivated by several factors, most notably the search for metals and precious stones (the first deposits were only located in the 1690s), agricultural expansion, and the capture of natives for slave labor. São Paulo to the South and Maranhão and Pará to the North were two of the main points from which the expansion movement grew. This was powered mostly, but not exclusively, by "the *mamelucos*, who regionally allied themselves to certain indigenous nations to benefit from their knowledge of the territory" (Andreazza/Nadalin 2011, 62–63).

Different expeditions came to investigate the territory and consequently increased contact with speakers of languages belonging mostly to the Macro-Ge trunk, which includes most of the indigenous languages of Brazil's *sertão* hinterlands. The first grammar of a Macro-Ge language spoken in the interior of the Northeast was published in 1699, *Arte de gramática da língua brasílica da nação cariri* ('The art of the grammar of the Brazilian language of the Cariri nation'), by the Jesuit Luís Vicêncio Mamiami. This publication confirms the hypothesis that "the phenomenon of the emergence of a 'general language' [in the sense of a vehicular language] occurred in more than one point across Brazil" (Houaiss 1985, 37); the two documented cases of indigenous languages in the colonial period were Tupinambá and Cariri.

Now is the time to ask which languages accompanied the explorers who set out from São Paulo and Maranhão/Pará. This question can be answered beginning with Rodrigues (1994, [1]1986; 1996): it was not the language of the colonizer. Along with the explorers, two "base" indigenous languages were spread: the *Paulista general language* and the *Amazonian general language*. From São Paulo, the Paulista general language expanded to Minas Gerais, the South of Goiás, Mato Grosso, and the North of Paraná. From Maranhão/Pará, the Amazonian general language advanced through the Brazilian Amazon; moving up the Rio Negro, it even reached the Venezuelan and Colombian Amazon.

Lingua geral and *lengua general* were expressions used by the Portuguese and Spanish colonists, literally meaning 'general language', to designate the languages of subjugated peoples spoken widely across the land, many of which surpassed their original geographic base and established communication between different ethnic and linguistic groups and inhabitants of different territories. Moving away from a more traditional historiography, which always uses the expression in the singular and often

8 Indigenous slavery was officially abolished in the State of Grão-Pará and Maranhão by the Law of June 6, 1755, and the provisions of this law were extended to the State of Brazil by the Charter of May 8, 1758.

only associated with "the most commonly language used along the Brazilian coast", Rodrigues proposed the diatopic, diachronic (1994, [1]1986), and linguistic (1996) distinction between two referents: one designating the Paulista general language, which originated in São Paulo in the 16th century and was "Tupiniquim-based", and another known as the Amazonian general language, originating in Maranhão/Pará in the 17th century and was "Tupinambá-based". The term "based" is used because both languages underwent changes as they were no longer exclusively spoken by Tupiniquim and Tupinambá natives and their initial *mameluco* descendants and were taken up by indigenous speakers of other languages, the Portuguese, and black people. In their respective coverage areas, these languages outstripped the language of the colonists: the Paulista language, for more than 250 years, and the Amazonian, for more than 300 years.[9] From the second half of the 19th century, just when it became a minority language due to the decimation of many of its speakers during the Cabanagem Revolt (1835–1840), the Amazonian general language came to be called *Nheengatu* ('good language'), and it is still spoken today in the Rio Negro valley.

Without distinguishing the various ethnic groups that comprised it, Wehling/Wehling (1994, 142) reported that, in the year 1700, the colonial population was 350,000 inhabitants. A rough picture of this population's linguistic reality can be drawn along the following lines:

> 'When the first gold mines were discovered in the *sertões* of *Minas Geraes* between 1693 and 1695, establishment of the Portuguese language in Brazil was still relatively precarious and reflected Portuguese colonization at the time: within the Brazilian territory, it occurred in a type of archipelago with its islands relatively isolated from one other. Concentrated in the two major centers of the colonial economy (the provinces of Bahia and Pernambuco), the Portuguese language faced a strong rival in this language known as *General*, which in different varieties predominated in the provinces of São Paulo and Maranhão, which during the colonial efforts of the time occupied a marginal position. In addition, the plantation, which marked the socio-economic structure of Bahia and Pernambuco, favored the emergence of pidginized and Creole varieties of Portuguese, in addition to the use of general African languages among the enslaved people that comprised the majority of the workforce in this agro-exporting economy'[10] (Lucchesi 2006, 351, our translation).

9 For discussions of the disappearance/decline of these languages, see Vitral (2001) and Bessa Freire (2004); for an in-depth introduction to the complex issue of general languages, see Nobre (2011).

10 "Quando as primeiras minas de ouro foram descobertas nos sertões das *Minas Geraes*, entre 1693 e 1695, a implantação da língua portuguesa no Brasil ainda era relativamente precária e refletia o traçado da própria colonização portuguesa na época, que configurava, no território brasileiro, uma espécie de arquipélago com ilhas relativamente isoladas umas das outras. Concentrada nos dois grandes centros dinâmicos da economia colonial—as Províncias da Bahia e de Pernambuco –, a língua portuguesa enfrentava uma forte rivalidade da língua chamada *geral*, que em matizes diferentes predominava nas Províncias de São Paulo e do Maranhão, as quais ocupavam uma posição marginal dentro do projeto colonial da época. Além disso, o caráter de *plantation*, que marcava a estrutura sócio-econômica da Bahia e de Pernambuco, favorecia a emergência de variedades pidginizadas e crioulizadas do português, além do uso de línguas gerais africanas, entre os escravos que formavam o grosso da mão de obra do empreendimento agro-exportador."

3 From *brasílicos* to *brasileiros*: the emergence of a proto-nation and a proto-national language

The 18th century is a watershed, not only from the point of view of consolidating Portuguese colonialism, but also from the perspective of the linguistic history of Brazil. Alencastro (2009, 17–18) proposed a distinction between the terms *brasílico* and *brasileiro* as a way of capturing the changes which were taking place at the time. *Brasílico* covered the majority of the inhabitants of the relatively isolated islands in the archipelago that was Portuguese colonial society in the Americas in the 16th and 17th centuries and the first half of the 18th century, who already differed from the king's subjects but did not yet possess the perception of belonging to a "proto-national" community. Beginning in the second half of the 18th century, *brasileiro* came to describe individuals who, because of Brazil's first great internal migration, which was driven by mining, were gradually becoming aware of their affiliation to supra-regional community and culture, speakers of the same language and residents of the same territory. Politically, examples of this consciousness can be seen in the two main separatist movements that occurred in the late 18th century, the Mineira Conjuration (1789–1792) and the Bahian Conjuration (1796–1799).

Understanding the history of the 1700s, and the linguistic developments which took place at that time, requires this era to be separated into two periods: the first, from 1690 to 1750, corresponds to the peak of mining, and a second started in the 1760s and 1770s, when gold production declined. This second period coincides with growing power on the part of Sebastião José de Carvalho e Melo, the future Marquis of Pombal (1699–1782). He became a plenipotentiary envoy during the reign of Dom José I (1750–1777) and was primarily responsible for a set of modernizing changes, the so-called Pombaline reforms, which were considered to be inspired by the Enlightenment and affected different areas of Portuguese life in Europe and overseas.

The discovery of gold not only sparked the first great internal migration, creating contact between different embryonic diatopic varieties of BP, but also impacted Portuguese migratory currents, which came in their largest waves and flowed from all regions of the mother country. In 1700, the Portuguese population was around 2,000,000 people and, throughout the 18th century, about 400,000 departed for Brazil. This fact, allied with natural growth, multiplied the white population tenfold over the course of the century (Marcílio 1999, 323–324).

The discovery of gold also impacted greatly on the slave trade, as the 18th century was the period with the most substantial importation of enslaved black African: from 1580 to 1690, 667,778 Africans entered Brazil; from 1691 to 1750, 1,034,171; from 1751 to 1808, 1,371,489, and finally from 1808 to 1850, 1,831,648 (Almeida 2014, 353–355). The slave trade was predominantly conducted in Congo and Angola, but large numbers of enslaved people were also brought from West Africa, especially the Bay of Benin; these were speakers of closely related Ewe-Fon languages. Their significant concen-

tration in the main mining city of that time, Vila Rica de Ouro Preto, led to the emergence of another vehicular language, recorded in 1731/1741 by Antônio da Costa Peixoto in *Obra nova da língua geral de mina* ('New work of the general mining language') (Castro 2002). It should also be noted that, because of inter-ethnic wars at the end of the 18th century, Salvador received a large contingent of people from what is currently Nigeria and that, since that time, this region has exhibited more examples of Yoruba culture, especially in the areas of religion and food (Castro 2001)

The question that must now be discussed is: why did the Portuguese language begin to spread widely across the Brazilian territory in the second half of the 18th century, instead of the Paulista general language, the Amazonian general language (both of which were disseminated widely), or a vehicular African language? The answers to this question have mobilized different lines of argument that are not mutually exclusive, although a hierarchy may be established between them.

Firstly, there is a tradition of historians of the Portuguese language that attributes the massive arrival of Portuguese, seduced by the discovery of gold and diamond mines, and above all, the Pombaline policies in the middle of the 1700s, greater responsibility for the "victory" of the colonizer's language (cf., for example, Teyssier 1997; Castro 1992). With regard to these policies, attention should be given to: 1) The Pombaline Directory, which ended Jesuit missionary work in indigenous villages and elevated them to the status of villages to be administered by a Director. The Director's main obligation was to "civilize the Indigenous", first and foremost by rooting out their languages of origin, indigenous languages or especially the "general language". The Directory, dating back to the year 1757, initially applied to villages in the State of Grão-Pará and Maranhão. In 1758, a charter extended it to the State of Brazil.[11] 2) The closing of the Jesuit colleges (by the Charter of June 28, 1759), followed by their expulsion from Portuguese territories (through the Law of September 3, 1759), established the system of "regal classes" for lower levels of study, corresponding to primary and secondary education. At the secondary level, where classes were independently configured, emphasis was placed on the humanities (Latin, Greek, and Hebrew grammar and rhetoric, as well as other subjects which were added later). At the primary level, one of the most significant innovations is found: the requirement to begin grammatical studies based on the Portuguese language, not Latin (Saviani 2013b).

Although long, the passage below, which was extracted from the Directory document, deserves mention:

> 'It has always been the unalterable maxim practiced in all Nations that conquered new Domains to immediately introduce their own language to the conquered Peoples, because it is indisputable that this is one of the most effective means of banishing the barbarity of these rustic People's ancient customs; and experience has shown that, while introducing them to the language of the

11 In 1798, the Directory was revoked, leading to the emancipation of local natives, on an equal footing with the other free inhabitants.

> Prince who conquered them, affection, veneration, and obedience to this same Prince is also established. Whereas all the educated nations of the world adhere to this prudent and solid system, in this instance of colonization the precise opposite was practiced, whereby the first colonizers merely saw it fit to establish the use of the language which they referred to as *geral* (i.e. the *língua geral* contact language), a truly abominable and devilish invention, so that, deprived of all those means that could civilize them, the Indigenous remained in the state of rustic and barbarous subjection in which we find them today. To banish this most pernicious abuse, it will be one of the main tasks of the Directors to establish the use of the Portuguese language, not permitting the boys and girls who belong to the schools and those Indigenous who are capable of receiving instruction in this subject to, by any means, use their own Nation's Language or the so-called general language (*língua geral*), but only Portuguese in the form that His Majesty has recommended in repeated orders, which until now have not been followed and have resulted in the total Spiritual and Temporal ruin of the State'[12] (in: Almeida 1997, 377–378, our translation).

A second line of argument is held by authors who defend the thesis that it was the Africans and their descendants who were mainly responsible, not only for the diffusion of Portuguese in the American colony, but also for the "formatting" of its demographically majority social variant, the so-called "popular" Brazilian Portuguese (cf., for example, Castro 2001; Ribeiro 1995; Mattos e Silva 2000). This point of view emphasizes that the population of Africans and Afro-Brazilians has always been extremely significant in the dynamics of colonial and post-colonial Brazilian demography. This group was present on the major economic fronts, not only through imports, but also migrating to various regions of Brazil due to domestic trafficking. Furthermore, they played multiple small-but-essential roles within the families of settlers and also in the extra-domestic space, both in the developing urban centers as well as the rural areas. Unable to establish their languages in Brazil, the Africans adopted the Portuguese of the colonizers and restructured it because, generally speaking, they had acquired the language as adults and were affected by the imperfect acquisition of a second language, and spread it throughout the Brazilian territory. As for explaining why

---

12 "*Sempre foi máxima inalteravelmente praticada em todas as Nações* que conquistaram novos Domínios *introduzir logo nos Povos conquistados o seu próprio idioma,* por ser indisputável que este é um dos meios mais eficazes para desterrar dos Povos rústicos a barbaridade dos seus antigos costumes; e ter mostrado a experiência que, ao mesmo passo que se introduz neles o uso da língua do Príncipe que os conquistou, se lhes radica também o afeto, a veneração e a obediência ao mesmo Príncipe. Observando, pois, todas as Nações polidas do Mundo este prudente e sólido sistema, nesta Conquista se praticou tanto pelo contrário, que só cuidaram os primeiros Conquistadores estabelecer nela o uso da Língua que chamaram geral, invenção verdadeiramente abominável e diabólica, para que, privados os Índios de todos aqueles meios que os podiam civilizar, permanecessem na rústica e bárbara sujeição, em que até agora se conservavam. Para desterrar este perniciosíssimo abuso, será um dos principais cuidados dos Diretores estabelecer nas suas respetivas Povoações o uso da Língua Portuguesa, não consentindo, por modo algum, que os Meninos e Meninas que pertencerem às Escolas e todos aqueles Índios que forem capazes de instrução nesta matéria usem da Língua própria das suas Nações ou da chamada geral, mas unicamente da Portuguesa, na forma que Sua Majestade tem recomendado em repetidas ordens, que até agora se não observaram com total ruína Espiritual e Temporal do Estado."

no African language survived at a spoken level, the dehumanizing politics of the slave trade— which began in Africa and separated persons of the same ethnicity and language,[13] to prevent them from reacting against the barbarous slave system (Houaiss 1985; Mattoso 1990; Mattos e Silva 2000)— and the impediments to enslaved people constituting stable, complete families (Mussa 1991) have been indicated as the main factors inhibiting the formation of sustainable nuclei of African languages. A final factor can be added, which certainly also prevented the establishment of such nuclei[14]: the adverse demographics observed among the captives, from premature deaths to low birth rates, which led to a negative growth rate, and the constant importation of slave labor from Africa (Schwarcz/Starling 2015, 79).[15]

With Pombal began secular public education, which differed from the previous period in which public education was expressly religious in nature. This was a period of regression because the dismantling of the Jesuit school system did not lead to an alternate and functional system of education. Saviani (2013a, 121) indicates the main reasons for the failure of the Pombaline educational reforms: the shortage of secular teachers, the fact that the training of these secular teachers was marked by Jesuit pedagogy, the colony's lack of a tax-collecting structure capable of financing the Royal classes, the political backlash when D. Maria I took the throne in 1777 after the death of D. José I, and finally, the fear that education would spread emancipationist ideas. So, assuming that to be successful, the Pombaline policy of imposing Portuguese should have been implemented through an effective school system, it is not difficult to conclude that the Pombaline Directory did have a broad reach, but certainly this was less broad than initially expected. This is seen in the fact that, in the State of Grão-Pará and Maranhão, the Amazonian general language was spoken through the mid-19th century. The large-scale arrival of the Portuguese in the 18th century also cannot be exclusively credited for the "victory" of the Portuguese language, since there was an equally large-scale arrival of Africans: the ratio of 30 % "whites" and 70 % "non-whites" remained stable throughout the colonial population, as seen below.

13 It was not always possible to separate co-linguals, due to the possibility of intercommunication between speakers of different African languages, especially those among the Bantu group.

14 Exceptions were the *quilombos*, communities comprised of escaped enslaved people seeking freedom, which formed between the 17th and the 19th centuries. Given the limits of this text, this important issue will not be discussed.

15 See Castro (2001, 76–78), which in addition to the socio-historical arguments, discusses the similarities between the language systems of the Bantu and Kwa languages and of Portuguese, as an unfavorable factor for the permanence of African languages and also the emergence of creoles in Brazil.

The second half of the 18th century brings with it the beginning of the proto-statistical stage of demographic studies in Brazil.[16] In theory, therefore, the table below would have "fewer rough estimates and be less indicative in nature" than the views for the 16th century:

Table 3: Racial composition in Brazil at the end of the colonial period (source: Alden 1997, 535, adapted for presentation).

Local	Whites	Mulattos and black people		Indigenous	Total
		Free persons	Slave		
Pará			23 %	20 %	80.000 (3.9 %)
Maranhão	31 %	17,3 %	46 %	5 %	78.860 – (3.8 %)
Piauí	21,8 %	18,4 %	36,2 %	23,6 %	58.962 – (2.9 %)
Pernambuco	28,5 %	42 %	26,2 %	3,2 %	391.986 – (19.1 %)
Bahia	19,8 %	31,6 %	47,0 %	1,5 %	359.437 – (17.5 %)
Mato Grosso	15,8 %			3,8 %	26.836 – (1.3 %)
Goiás	12,5 %	36,2 %	46,2 %	5,2 %	55.422 – (2.7 %)
Minas Gerais	23,6 %	33,7 %	40,9 %	1,8 %	494.759 – (24.1 %)
Rio de Janeiro	33,6 %	18,4 %	45,9 %	2 %	229.582 – (11.2 %)
São Paulo	56 %	25 %	16 %	3 %	208.807 – (10.2 %)
Rio Grande do Sul	40,4 %	21 %	5,5 %	34 %	66.420 – (3.2 %)
Average for 8 jurisdictions[17]	28,0 %	27,8 %	38,1 %	5,7 %	

At the end of the colonial period, the Brazilian population reached 2,051,071 inhabitants, still quite unevenly distributed. Its growth and changes in geographical distribution are notable compared to the 1600s and 1700s. With more than half concentrated in the Central-South, one of the factors that led to this change in demographic density was the relocation of the capital from Salvador to Rio de Janeiro in 1763. In general, an average of 28.0 % of inhabitants were white, 65.9 % were of black African descent and mixed-race, distributed among enslaved and free persons, and 5.7 % were Natives. The contrast between approximately 30 % "whites" and 70 % "non-

16 There are no general data about the social diffusion of writing. Without attributing sources, Houaiss (1985, 88–89) states that from the 16th to the 18th centuries, the percentage of educated, literate Portuguese varied between 0.5 % and 1 % of the population.

17 With the exception of Mato Grosso, Pará and Rio Grande do Sul, whose data are incomplete/lacking.

whites", which is seen at the end of the 16th century, continues. What stands out, however, is that black African people and their *mulatto* descendants predominate among "non-whites", in comparison with Natives and their descendants. Furthermore, even among the so-called "whites", there was surely a significant number of mixed-race people (now, predominantly *mulatto*) who had undergone the social process of "whitening". It should be noted that the population of *mulattos* and free black people is already significant—27.8 % of the total, a factor that can also be considered favorable to the dissemination of the Portuguese language, the language of the dominant society.

4 Independence, urbanization and social stratification of BP

The present section is dedicated to the 19th, 20th, and early 21st centuries, focusing on understanding the major processes of change associated with the socio-linguistic configuration of modern Brazil.

4.1 The 19th century

Because of the Napoleonic invasions of Portugal, the royal family and the court of Lisbon moved to Brazil, and D. João VI's reign in the Americas lasted from 1808 to 1821, when relations with the mother country changed profoundly. The colony's elevated political status—specifically the State of Brazil—became more explicit in 1815, with the Constitution of the United Kingdom of Portugal, Brazil and the Algarves. Therefore, it can be said that Independence in 1822 was more of a transition than a great rupture. Regardless, between 1822 and 1889, Brazil was formed as a monarchic nation-state, the Empire of Brazil, when the Republic was declared.

It is estimated that 10,000 to 15,000 Portuguese followed the royal family and that, for the thirteen years they were in Rio de Janeiro, the city's population doubled from 50,000 to 100,000 inhabitants (Fausto 1995, 121, 125). Consequently, it is possible to speak of a "re-lusitaniation" of the city, which upon being designated the capital of the very Portuguese empire, became the space where the linguistic norms of higher social prestige were incubated, norms which were molded by the presence of the highest strata of Portuguese society. The question, however, is more complex, as the quote below clearly reveals:

> '[...] the crass and solecistic language of D. Pedro I [son of D. João VI], who arrived here in 1808, has nothing to do with the Brazilian environment itself. His Highness, not having had serious studies, was deprived along with the colonial rabble. [...] The royal speech contained so many atroci-

> ties that his father admonished him in 1824: "When you write, don't forget that you are a Prince and that your writings are seen all over the world and you must be careful, not only in what you say, but also in the way you explain"'[18] (Luís Norton, *A corte de Portugal no Brasil*, 1938, in: Silva Neto 1986, 79, our translation).

Judging from the meaning of the Portuguese word *solecismo*, which involves "meddling, in the norms of a language using random syntactical constructions, usually by people who do not have complete dominance of its rules (for example, so-called errors of concordance, verb usage, collocation, poor construction of a compound period, etc.)" (Houaiss/Villar 2001, 2601), the speech and writing of the person, who in 1824 became the first emperor of Brazil, were contaminated with syntactical constructions against which the Brazilian purist grammatical tradition would fight beginning in the second half of the 19th century. This quote is exemplary in demonstrating the impact of socio-linguistic changes defined as "bottom-up" on conformation with linguistic standards in the speech of the highest socio-economic, and even political strata, in Brazil.

Nevertheless, it is impossible to refute the scenario of linguistic "duality" outlined by Silva Neto (1986, 80) in the following passage:

> '[...] from the beginning of colonization until 1808, and from then onward with increasing intensity, the linguistic duality was notable between the social "cream of the crop", the nursery of the white and mixed-race people who rose socially, and the plebeians, the descendants of Natives, black and mixed-race people in the colony.'[19]

This framework, according to Lucchesi (2015), only began to be modified with the 1930 revolution that put an end to the Old Republic (1889–1930), which was fundamentally still a rural country, an exporter of agricultural commodities, marked by the over-exploitation of manual labor and the reduced internal consumer market. This revolution started an accelerated process of industrialization and urbanization in Brazilian society.

Notable among the significant transformations of Dom João's reign (including in its cultural plan the installation of the press, the foundation of higher education, the creation of the Royal Library, and the promotion of artistic missions from abroad) is the establishment of settlement policies involving European immigrants, which per-

18 "[...] a grosseira e solecística linguagem de D. Pedro I [filho de D. João VI], aqui chegado, em 1808, nada tem que ver com o ambiente brasileiro em si. É que Sua Alteza, sobre não ter estudos sérios, privava com a ralé colonial. [...] Eram tantas as barbaridades da fala real que, em 1824, assim o admoestava o pai: 'Quando escreveres, lembra-te de que és um Príncipe e que os teus escritos são vistos por todo o mundo e deves ter cautela não só no que dizes, mas também no modo de te explicares'."

19 "[...] dos princípios da colonização até 1808, e daí por diante com intensidade cada vez maior, se notava a dualidade linguística entre a nata social, viveiro de brancos e mestiços que ascenderam socialmente, e a plebe, descendente dos índios, negros e mestiços da colônia."

mitted the entry of new actors onto the Brazilian linguistic scene: the languages of immigration. In 1818, the first immigrants settled in Petrópolis and Nova Friburgo (in the Province of Rio de Janeiro) and Leopoldina and São Jorge dos Ilhéus (in the Province of Bahia). On the colonial coffee plantations of Leopoldina, where slave labor was used on a large scale and where the settlers were German-speaking Swiss and French, a creolized variety of Portuguese (the Helvécia dialect) emerged and was registered in the 1860s (Ferreira 1984); certainly this was only the visible tip of a broad and deep iceberg.

After independence, incentives to immigration were broadened. The intention was to occupy empty and/or disputed areas, to promote the whitening of the population, and above all, to replace slave labor with salaried European labor (and later, Asian labor). This was associated with the idea of economic progress, given the belief in these groups' ethnic and cultural superiority. From 1840, coffee, the new great economic catalyst, became the primary export product. This fact, along with the official end of the slave trade in 1850, the abolition of slavery in 1888, and the strong industrialization of the country in the 20th century, greatly boosted foreign immigration. It is estimated that, from 1818 to 1970, over 5,500,000 immigrants entered the country. It is also estimated that, of the 4,500,000 who arrived between 1850 and 1918, 2,500,000 moved to the state of São Paulo (Andreazza/Nadalin 2011, 66–69), which became Brazil's largest economic center. The immigrants and their languages were not distributed homogeneously throughout the country. They were present in the South and Southeast, predominantly in rural areas, such as large coffee estates or small farms, as well as urban areas, especially the city of São Paulo. In many areas, they formed settlements and had little contact with the national population. Consequently, they tended to maintain their ethnic and cultural characteristics, notably their languages (Kreutz 2000).

The following data come from the 1872 Census, which opens the age of statistical demographic analysis in Brazil:

Table 4: Population distribution in relation to sex, servile state, and race (source: IBGE, Electoral census of 1872).

Sex	Race	Population		Total
		Free	Slave	
Male	Whites	1.971.772 (20 %)		1.971.772 (20 %)
	Pardos	1.673.971 (17 %)	252.824 (2 %)	1.926.795 (19 %)
	Black people	472.008 (5 %)	552.346 (6 %)	1.024.354 (11 %)
	Caboclos	200.948 (2 %)		200.948 (2 %)
Female	Whites	1.815.517 (18 %)		1.815.517 (18 %)
	Pardos	1.650.307 (17 %)	224.680 (2 %)	1.874.987 (19 %)

Table 4: (continued)

Sex	Race	Population		Total
		Free	Slave	
	Black people	449.142 (4 %)	480.956 (5 %)	930.098 (9 %)
	Caboclos	186.007 (2 %)		186.007 (2 %)
Total		8.419.672 (85 %)	1.510.806 (15 %)	9.930.478 (100 %)

In 1872, the Brazilian population reached 9,930,478 inhabitants, 94.1 % of whom were in rural areas. Whites and mixed-race people, known as *pardos* (i.e. *mulattos*), were equal in number, with each comprising 38 % of the total. Next were black people, representing 20 %. The decrease in this group may be attributed to the official end of the slave trade in 1850, or to racial mixing itself. Finally, at a considerably lower level, were *caboclos* at 4.0 %. Since *caboclo* is a general designation for mixed-race Natives, the Natives are "invisible" in the 1872 census. The contrast between "whites" and "non-whites" began to fade due to the entry of European immigrants.

Beginning in the second half of the 19th century, independent Brazil was the stage for discussions about the process of linguistic standardization, with developments in two spheres. One features heated controversy among writers on the "issue of the Brazilian language"; the other involves initial production of grammar by authors born in Brazil. While romantic writers sought to affirm the new nation's identity based on a literary language that could be called "national", grammarians and even a significant group of intellectuals defended total subservience to the standards of the former colonial motherland, which obviously was indifferent to the solecisms contained within the writing and speech of Brazil's first emperor. In other words, "the [grammatical] encoding that occurred here in the second half of the 19th century did not take the educated/common/*standard* (...) Brazilian norms of that time as a reference" (Faraco 2008, 80).

What did a snapshot of the social distribution of writing in this newly independent country look like, in a place which had begun to discuss the process of linguistic standardization?

Table 5: Population distribution in relation to sex, servile state, and level of education (source: IBGE, Electoral census of 1872).

Servile state	Sex	Level of education		Total
		Literates	Illiterates	
Free	Male	1.012.097 – 23,4 %	3.306.602 – 76,6 %	4.318.699 – 100 %
	Female	550.981 – 13.4 %	3.549.992 – 86.6 %	4.100.973 – 100 %
	Total	1.563.078 – 19 %	6.856.594 – 81 %	8.419.672 – 100 %
Slave	Male	958 – 0,1 %	804.212 – 99.9 %	805.170 – 100 %
	Female	445 – 0,1 %	705.191 – 99.9 %	705.636 – 100 %
	Total	1.403 – 0,1 %	1.509.403 – 99,9 %	1.510.806 – 100 %

In 1872, Brazil was predominantly rural, and its population was predominantly illiterate: 76.6 % of free men were illiterate, as were 86.6 % of free women, and illiteracy was practically universal among enslaved persons, reaching 99.9 % among both men and women.

The royal class system instituted by Pombal remained until the Additional Act of the Imperial Constitution, in 1834, which transferred responsibility for primary and secondary education from the central government to the provincial governments and, in the process, failed to promote the creation of a national system of public instruction at these two levels (Saviani 2013a, 122–124), a situation which remains practically unchanged to the present day. Higher education and education in the capital fell to the central government. The rates of illiteracy mentioned above fundamentally demonstrate two aspects: the ban on enslaved persons attending schools, which lasted until 1888, and the paltry investment in education by the Brazilian Empire. During the reign of Dom Pedro II (the son of Pedro I), which lasted from 1840 to 1889, "the annual average of financial resources invested in education was 1.80 % of the imperial government's budget, and an average of 0.47 % was earmarked for primary and secondary education" (Saviani 2006, 28–29).

4.2 The 20th and early 21st centuries

The 20th century hosts the two main processes of transformation associated with Brazil's current sociolinguistic configuration: the country's change from predominantly rural-agrarian to urban-industrial, and with regard to urbanization and industrialization, growth in the population's education and literacy rates.

The following data illustrate the country's population growth and process of urbanization:

Table 6: Distribution of the rural and urban populations in Brazil from 1872 to 2010 (source: IBGE, Demographic Census, in Lucchesi 2015, 144).

Year	Total	Urban population		Rural population	
1872	9.930.478	582.749	5,9 %	9.347.729	94,1 %
1890	14.333.915	976.038	6,8 %	13.357.877	93,2 %
1900	17.438.434	1.644.149	9,4 %	15.784.285	90,6 %
1920	30.635.605	3.287.448	10,7 %	27.348.157	89,3 %
1940	41.326.315	12.880.182	31,2 %	28.448.133	68,8 %
1950	21.944.397	18.782.891	36,2 %	33.161.506	63,8 %
1960	70.992.343	32.004.817	45,1 %	38.987.526	54,9 %
1970	94.508.583	52.904.744	56,0 %	41.603.839	44,0 %
1980	121.150.573	82.013.375	67,7 %	39.137.198	32,3 %
1990	146.917.459	110.875.826	75,5 %	36.041.633	24,5 %
2000	169.590.693	137.755.550	81,2 %	31.835.143	18,8 %
2010	190.755.799	160.925.792	84,4 %	29.830.007	15,6 %

The majority presence of Brazilians in cities is a very recent phenomenon. It became perceptible in the 1940s and 1950s, as a result of the industrialization process that began in 1930. It peaked at 31.3 % and 36.2 %, respectively, and was consolidated in the 1970s and 1980s, mainly because of the rural population's exodus to industrialized urban centers. Today, approximately 85 % of Brazilians live in urban centers. As noted by Bortoni-Ricardo (2011, [1]1985) in her study of the birth of a dialect in Brazlândia, a satellite city of Brasilia, which became Brazil's new capital in 1960, the first fundamental characteristic for understanding contemporary BP is the transformation of rural language varieties (which are characterized, according to this author, by a surprisingly high degree of uniformity) into non-standard urban varieties. She designates these as *rurbanas* ('rurban'), which on the dialectical *continuum* are adjacent to isolated and highly stigmatized rural varieties. They are spoken not only in the cities by illiterate or semi-literate individuals of the lower classes with mostly rural backgrounds, but also in rural areas, where speakers experience the technological and modernizing influence of the cities.

This, however, is not the only aspect to consider. Bortoni-Ricardo (2002, [1]1998) proposes a model comprised of three *continua* to linguistically characterize contemporary Brazil[20]: the previously-mentioned rural-urban *continuum*, which is aimed at

20 Cf. Lucchesi (2015) for discussion of the complementarity between his proposed sociolinguistic polarization model and the model of the three continua.

analysis of the sociological aspects of the speaker; the oral-literate *continuum*, referring to the analysis of the literate social practices in which the individual takes part; and the stylistic monitoring *continuum*, which is related to analysis of the psychological processes of planning and attention at the moment of speaking. The data below allow for reflections on changes related to the oral-literate *continuum*:

Table 7: Illiteracy rate in Brazil from 1920 to 2010, among the population over 15 years of age (source: IBGE, Demographic Census, in Lucchesi 2015, 149).

Year	Total	Illiterates	%
1920	17.557.282	11.401.715	64,9 %
1940	23.709.769	13.242.172	55,9 %
1950	30.249.423	15.272.632	50,5 %
1960	40.278.602	15.964.852	39,6 %
1970	54.008.604	18.146.977	33,6 %
1980	73.542.003	18.716.847	25,5 %
1991	95.810.615	18.587.446	19,4 %
2000	119.533.048	16.294.889	13,6 %
2010	144.823.504	13.949.729	9,6 %

It can be noted that the near-eradication of illiteracy in Brazil is a fairly recent phenomenon: until the 1950s, half of the population was illiterate. Progressive access to schooling by individuals originally located in "rurban" settings appears as a factor for the second fundamental characteristic of contemporary BP, that is, a set of changes that may take place in a "top-down" manner by exposing speakers to the linguistic standards prized by the literate culture. However, as the advance in the oral-literate *continuum* is still deeply uneven in the country, a complex set of gradient and interpenetrating norms can be identified among the poles of Brazil's contemporary sociolinguistic web (↗10 Sociolinguistics and L(↗3 The history of linguistic contact underlying Brazilian Portuguese).

5 Conclusions

Despite a history of glottocide, which resulted in predominant monolingualism, the linguistic ecology of Brazil in the early 21st century is abundant (↗11 Brazilian Portuguese: contemporary language contacts). In 2010, the country reached a population of 190,755,799 inhabitants. Among these, 896,900 are Natives; 274 indigenous languages were recorded, and were spoken by 37.4 % of the Natives (cf. the 2010 Census).

Among the minority languages are approximately 30 immigrant languages, two creole languages (Karipuna and Galibi-Marworno), as well as two sign languages, Libras and Kaapor (Morello 2012, 13–14). Of the 200 to 300 African languages that crossed the Atlantic (Petter 2006), none remain spoken, but traces of their presence remain in religious rituals and even the so-called "secret languages" derived from Bantu-based vehicular languages (Vogt/Fry 1996; Queiroz 1998).

Besides Portuguese and Libras,[21] no other language has the same national, or even regional, coverage. There is, however, a significant group of municipalities where the inhabitants predominantly speak indigenous or immigrant languages, which make them "potential instances for language management" (Morello 2012, 12). Consequently, based on this vision, the municipality with the greatest number of languages in the Americas (São Gabriel da Cachoeira in Amazonas, where 23 languages are spoken) created a policy of officially recognizing multiple languages on a municipal level, an innovative initiative in the field of language policies in the country. This municipality recognized two indigenous languages as official, namely Baniwa and Tukano, and it also recognized Nheengatu, which is a metaphor for the linguistic social history of Brazil: a language of resistance, but also, in the words of Mattos e Silva (2004), "the living fruit of the death of many languages".

6 References

Alden, Dauril (1997), *O período final do Brasil Colônia, 1750–1808*, in: Leslie Bethel (ed.), *História da América Latina*, vol. 2: *América colonial*, São Paulo/Brasília, EDUSP/Fundação Alexandre de Gusmão, 527–592.

Alencastro, Luiz Felipe (2009), *Os africanos e as falas africanas no Brasil*, in: Charlotte Galves/Helder Garmes/Fernando Rosa Ribeiro (edd.), *África-Brasil: caminhos da língua portuguesa*, Campinas, Editora da UNICAMP, 15–26.

Almeida, Rita Heloísa de (1997), *O Diretório dos Índios: um projeto de "civilização" no Brasil do Século XVIII,* Brasília, Editora da UnB.

Almeida, Marcos Abreu Leitão de (2014), *Tráfico de africanos para o Brasil*, in: Ivana Stolze Lima/Laura do Carmo (edd.), *História social da língua nacional*, vol. 2: *Diáspora africana*, Rio de Janeiro, Nau/FAPERJ, 73–104.

Anchieta, José de (1980, [1]1595), *Arte de grammatica da lingoa mais usada na costa do Brasil*, Salvador, Universidade Federal da Bahia.

Andreazza, Maria Luíza/Nadalin, Sérgio Odilon (2011), *História da ocupação do Brasil*, in: Heliana Mello/Cléo Altenhofen/Tommaso Raso (edd.), *Os contatos linguísticos no Brasil*, Belo Horizonte, Editora da UFMG, 57–72.

Barros, Maria Cândida D. M./Borges, Luiz C./Meira, Márcio (1996), *A língua geral como identidade construída*, Revista de Antropologia 39/1, 191–219.

21 Law No. 10.436 of April 20, 2002 regulated the use of Libras and made Brazil officially bilingual in its entirety.

Bessa Freire, José Ribamar (2004), *Rio Babel: a história das línguas na Amazônia*, Rio de Janeiro, Atlântica.

Bortoni-Ricardo, Stella Maris (2002, [1]1998), *Um modelo para a análise sociolingüística do português do Brasil*, in: Marcos Bagno (ed.), *A linguística da norma*, São Paulo, Edições Loyola, 333–346.

Bortoni-Ricardo, Stella-Maris (2011, [1]1985), *Do campo para a cidade: estudo sociolinguístico de migração e redes sociais*, São Paulo, Parábola.

Castro, Ivo (1992), *História da língua portuguesa. Relatório (Programa e manual da cadeira*), Faculdade de Letras, Universidade de Lisboa.

Castro, Yeda Pessoa de (1983), *Das línguas africanas ao português brasileiro*, Afro-Ásia 14, 81–101.

Castro, Yeda Pessoa de (2001), *Falares africanos na Bahia: um vocabulário afro-brasileiro,* Rio de Janeiro, Academia Brasileira de Letras/Topbooks.

Castro, Yeda Pessoa de (2002), *A língua mina-jeje no Brasil: um falar africano em Ouro Preto do século XVIII*, Belo Horizonte, Fundação João Pinheiro/Secretaria da Cultura do Estado de Minas Gerais.

Cintra, Jorge Pimentel (2013), *Reconstruindo o mapa das Capitanias Hereditárias*, Anais do Museu Paulista, 21/2, 11–45.

Couto, Jorge (1998), *A construção do Brasil: ameríndios, portugueses e africanos, do início do povoamento a finais de Quinhentos*, Lisboa, Edições Cosmos.

Dias, Pedro (1697), *Arte da Lingua de Angola offerecida a virgem senhora N. Do Rosario, Mãy, & Senhora dos mesmos Pretos, pelo P. Pedro Dias, da Companhia de Jesu*, Lisboa, Officina de Miguel Deslandes, Impressor de Sua Magestade.

Faraco, Carlos Alberto (2008), *Norma culta brasileira: desatando alguns nós*, São Paulo, Parábola.

Fausto, Boris (1995), *História da Brasil*, São Paulo, EDUSP/FDE.

Fausto, Carlos (1992), *Fragmentos de história e cultura tupinambá: da etnologia como instrumento crítico de conhecimento etno-histórico*, in: Manuela Carneiro da Cunha (ed.), *História dos índios no Brasil*, São Paulo, FAPESP/Companhia das Letras/SMC, 381–396.

Ferreira, Carlota (1984), *Remanescentes de um falar crioulo brasileiro*, in: Carlota Ferreira et al. (edd.), *Diversidade do português do Brasil*, Salvador, EDUFBA, 21–32.

Gândavo, Pero de Magalhães de (1965, [1]1576), *Tratado da província do Brasil*, Rio de Janeiro, Instituto Nacional do Livro.

Greenberg, Joseph H. (1966), *The Languages of Africa*, Bloomington, Indiana University.

Grupo de Trabalho da Diversidade Linguística do Brasil (GTDL), Relatório de Atividades (2006–2007), http://portal.iphan.gov.br/uploads/ckfinder/arquivos/12_2%20Audi%C3%AAncia%20P%C3%BAblica%20Debate%20Diversidade%20Lingu%C3%ADstica%20no%20Brasil.pdf, last accessed 13.02.2022.

Houaiss, Antônio (1985), *A língua portuguesa no Brasil*, Rio de Janeiro, UNIBRADE.

Houaiss, Antônio/Villar, Mauro de Salles (2001), *Dicionário Houaiss da língua portuguesa*, Rio de Janeiro, Objetiva.

IBGE—Instituto Brasileiro de Geografia e Estatística, http://www.ibge.gov.br/home/, *last accessed 13.02.2022*.

IBGE—Instituto Brasileiro de Geografia e Estatística, https://biblioteca.ibge.gov.br/biblioteca-catalogo?id=225477&view=detalhes, *last accessed 13.02.2022*.

Kreutz, Lúcio (2000), *A educação de imigrantes no Brasil*, in: Eliane Lopes/Luciano Faria Filho/Cynthia Veiga (edd.), *500 anos de educação no Brasil*, Belo Horizonte, Autêntica, 347–370.

Lobo, Tânia/Oliveira, Klebson (2013), *Ainda aos olhos da Inquisição: novos dados sobre níveis de alfabetização na Bahia em finais de Quinhentos*, in: Rosario Álvarez et al. (edd.), *Ao sabor do texto. Estudos dedicados a Ivo Castro*, Santiago de Compostela, Universidade de Santiago de Compostela, Servizo de Publicacións e Intercambio Científico, 273–291.

Lucchesi, Dante (2006), *Século XVIII, o século da lusofonização do Brasil*, in: Werner Thielemann (ed.), *Século das luzes: Portugal e Espanha, o Brasil e a Região do Rio da Prata*, Frankfurt, TFM, 351–370.

Lucchesi, Dante (2015), *Língua e sociedade partidas: a polarização sociolinguística do Brasil*, São Paulo, Contexto.

Mamiami, Luiz Vincencio (1877, [1]1699), *Arte de Grammatica da Lingua Brasilica da Naçam Kiriri (com notas introdutórias de Batista Caetano de Almeida Nogueira)*, Rio de Janeiro, Bibliotheca Nacional.

Marcílio, Maria Luíza (1999), *A população brasileira colonial*, in: Leslie Bethel (ed.), *História da América Latina*, vol. 2: *América colonial*, São Paulo/Brasília, EDUSP/Fundação Alexandre de Gusmão, 311–338.

Mattos e Silva, Rosa Virgínia (2000), *Uma interpretação para a generalizada difusão da língua portuguesa no território brasileiro*, Gragoatá 9, 11–27.

Mattos e Silva, Rosa Virgínia (2004), *Ensaios para uma sócio-história do português brasileiro,* São Paulo, Parábola

Mattoso, Kátia de Queirós (1990), *Ser escravo no Brasil*, São Paulo, Editora Brasiliense.

Morello, Rosângela (2012), *Política de cooficialização de línguas no Brasil*, Revista do Instituto Internacional da Língua Portuguesa 1, Colóquio Internacional de Maputo, https://www.yumpu.com/pt/document/read/62885013/plato-vol-1-n-1-2012-a-diversidade-linguistica-nos-paises-da-cplp, last accessed 13.02.2022.

Mussa, Alberto (1991), *O papel das línguas africanas na história do português do Brasil*, Master's thesis, Rio de Janeiro, Faculdade de Letras da Universidade Federal do Rio de Janeiro.

Nobre, Wagner Carvalho de Argolo (2011), *Introdução à história das línguas gerais no Brasil: processos distintos de formação no período colonial*, Master's thesis, Salvador da Bahia, Instituto de Letras da Universidade Federal da Bahia.

Oliveira, Maria Inês Côrtes (1997), *Quem eram os "negros da Guiné"? A origem dos negros na Bahia*, Afro-Ásia 19/20, 19–73.

Petter, Margarida Maria Taddoni (2006), *Línguas africanas no Brasil*, in: Suzana Alice Marcelino Cardoso/Jacyra Andrade Mota/Rosa Virgínia Mattos e Silva (edd.), *Quinhentos anos de história linguística do Brasil*, Salvador, Secretaria da Cultura e Turismo do Estado da Bahia, 117–142.

Queiroz, Sônia (1998), *Pé preto no barro branco: a língua dos negros da Tabatinga*, Belo Horizonte, EDUFMG.

Ribeiro, Darcy (1995), *O povo brasileiro: a evolução e o sentido do Brasil*, São Paulo, Companhia das Letras.

Rodrigues, Aryon Dall'Igna (1993), *Línguas indígenas: 500 anos de descobertas e perdas*, D.E.L.T.A. 9/1, 83–103.

Rodrigues, Aryon Dall'Igna (1994, [1]1986), *Línguas brasileiras: para o conhecimento das línguas indígenas*, São Paulo, Edições Loyola.

Rodrigues, Aryon Dall'Igna (1996), *As línguas gerais sul-americanas*, Papia 4/2, 6–18.

Sartori, Ana (2016), *Pela pena do Santo Ofício: difusão social da escrita nas capitanias de Pernambuco, Itamaracá e Paraíba em finais Quinhentos*, Doctoral thesis, Salvador, Instituto de Letras da Universidade Federal da Bahia.

Saviani, Dermeval (2006), *O legado educacional do "breve século XIX" brasileiro*, in: Dermeval Saviani et al., *O legado educacional do século XIX, Campinas*, Autores Associados.

Saviani, Dermeval (2013a), *Aberturas para a história da educação: do debate teórico-metodológico no campo da história ao debate sobre a construção do sistema nacional de educação no Brasil*, Campinas, Autores Associados.

Saviani, Dermeval (2013b), *História das ideias pedagógicas no Brasil*, Campinas, Autores Associadas.

Schwarcz, Lilia Moritz/Starling, Heloisa (2015), *Brasil: uma biografia*, São Paulo, Companhia das Letras.
Silva Neto, Serafim da (1986, [1]1950), *Introdução ao estudo da língua portuguesa no Brasil*, Rio de Janeiro, Presença/Instituto Nacional do Livro.
Teyssier, Paul (1997, [1]1982), *História da língua portuguesa*, São Paulo, Martins Fontes.
Vitral, Lorenzo (2001), *Língua geral versus língua portuguesa: a influência do processo civilizatório*, in: Rosa Virgínia Mattos e Silva (ed.), *Para a história do português brasileiro: primeiros estudos*, São Paulo, Humanitas, 303–315.
Vogt, Carlos/Fry, Peter (1996), *Cafundó: a África no Brasil*, São Paulo, Companhia das Letras.
Wehling, Arno /Wehling, Maria José C. de M. (1994), *Formação do Brasil colonial*, Rio de Janeiro, Nova Fronteira.

Dante Lucchesi and Alan Baxter

3 The history of linguistic contact underlying Brazilian Portuguese

Abstract: The current chapter provides a historical overview of how the contact between languages shaped the development of varieties of the Portuguese language in Brazil. Prior to sketching details of the historical linguistic contacts, an outline of the controversy surrounding this view is presented, showing how research of recent decades rejects the hypothesis of secular drift as a source of the massive variation inherent in Popular Brazilian Portuguese. The chapter then presents a brief historical overview of the contact between languages in Brazil, covering the major phases of language contact in the colonial period that yielded the sociolinguistic divide that typifies Brazilian society today. Subsequently, the concept of irregular linguistic transmission is introduced to account for how Brazil's historical situations of language contact altered the grammatical structure of Popular Brazilian Portuguese. The final sections of the chapter describe the morpho-syntactic characteristics of Brazilian Portuguese that stem from changes caused by language contact, while drawing parallels with the structuring processes that formed the Portuguese-lexified creole languages of Africa.

Keywords: Popular Brazilian Portuguese, language contact, irregular language transmission, creolization, secular drift

1 Introduction

Brazil's linguistic panorama is paradoxical. On the one hand, it is one of the countries with the highest linguistic homogeneity worldwide, because the overwhelming majority of its population is monolingual in Portuguese[1] (Massini-Cagliari 2004). On the other hand, there are 274 indigenous languages spoken, spread over five typologically different families, as well as languages resulting from immigration, comprising predo-

1 Figures published by Ethnologue (http://www.ethnologue.com/country/BR and www.silbrasil.org.br, last accessed 14.02.2022) report a total population of 204,735,000 (for 2017), with a population of immigrant language speakers of approximately 532,000, plus a population of indigenous language speakers of approximately 600,000. No statistics are available regarding the actual number of Portuguese-speaking monolinguals. However, based on the Ethnologue/SIL Brasil figures, we estimate that 98 % of the population are monolingual in Portuguese.

https://doi.org/10.1515/9783110405958-004

minantly varieties[2] of Italian, German, Japanese, Korean and Spanish. This situation is the result of the colonization process of Brazil, which began almost 500 years ago, when the Portuguese language was superimposed on many hundreds of indigenous Brazilian languages and African languages introduced by the slave trade. Such facts led Mattos e Silva (2004, 14) to define the linguistic history of Brazil as the transition from a widespread multilingualism to a localized multilingualism.

It is estimated that more than a thousand indigenous languages were spoken in Brazil when the Portuguese began to effectively colonize Brazil in 1532 (Rodrigues 1986). Most of these languages disappeared in the first century of colonization, along with the extermination of their speakers. To replace the indigenous workforce, the colonizers began to import slaves from Africa, already in the 16th century. It is estimated that between 1550 and approximately 1850, some four million enslaved Africans were imported, speaking over 200 languages often typologically quite different from each other (Petter 2006, 124). As such, Brazil's linguistic history is the story of a long and violent process of linguistic homogenization.

However, the imposition of Portuguese occurred in contexts that were less than favorable to language acquisition, and Amerindians and Africans came to use much altered second language varieties of Portuguese. The nativization of these second language varieties of Portuguese among the descendants of Amerindians and Africans failed to produce creole languages, unlike the situation in the Caribbean. Nevertheless, it triggered changes that determined the features that today differentiate the popular varieties from the language of the literate elite (Lucchesi 2009), causing the situation of sociolinguistic polarization that is now typical of modern Brazil (Lucchesi 2015a). In comparison with the language of educated urban speakers, the popular varieties of Brazilian Portuguese display a notable morphological reduction. This difference, resulting from a history of language contact and a process of *irregular linguistic transmission* (Lucchesi/Baxter 2009), characterizes a major sociolinguistic division in Brazilian society.

This view of the sociolinguistic history of Brazil has faced strong resistance within the Brazilian tradition of linguistic research, dominated by the notion that the historical development of a language is determined by the logic of its internal structure. Even such prominent sociolinguists as Naro/Scherre (2007), adhering to the concept of language drift proposed by Sapir (1921), argued that the morphological simplification of Brazilian Portuguese was the result of a built-in secular trend imported from Portugal. However, the controversy surrounding the origins of Brazilian Portuguese is being overcome with advances in historical and empirical research on Brazilian vernacular speech, especially in isolated rural communities formed by direct descen-

2 In this chapter, the term variety refers to a geographic and/or social variety or lect of a particular language.

dants of African slaves (Lucchesi/Baxter/Ribeiro 2009), and with advances in comparative research on African varieties of Portuguese.

This chapter will present a historical overview of how the contact between languages affected the development of varieties of the Portuguese language in Brazil. The first section will summarize the controversy concerning this issue. The second will trace a brief historical overview of the contact between languages in Brazil. The third will present the concept of *irregular linguistic transmission*, to explain how the grammatical structure is affected in the contact situations that occurred. Finally, the fourth and final section will describe the characteristics of Brazilian Portuguese that result from changes induced by contact between languages.

2 The question of the origins of Brazilian Portuguese

Until the 19th century, native speakers of Portuguese who were children of native speakers of Portuguese comprised only one-third of Brazil's population. The other two thirds were made up by Africans, Amerindians and their descendants, who spoke Portuguese as a second language, or a nativized variety thereof, in a multilingual context. Thus, the ethnic dimension of Brazil's linguistic divide was very clear to the more astute observers of the time, as can be seen in this passage from the entry *Elemento Negro* ('Black Element') in the *Dicionário Gramatical* ('Grammatical Dictionary') published by Ribeiro (1906, [1]1889):

> 'Under the entry *Black Element*, we group all kinds of alterations produced in the Brazilian language by the influence of African languages spoken in Brazil. These changes are not as superficial as some scholars argue: on the contrary, they are quite deep, not only in regard to vocabulary, but even in the grammar system of the language' (our translation).

The Portuguese philologist Adolfo Coelho (1967, [1]1880–1886, 43–44), a pioneer in the study of creole languages, also realized this and associated the Portuguese spoken by the masses in Brazil with the Portuguese 'creole dialects':

> 'Various characteristic features of Creole dialects recur in Brazil; such as the tendency for the elimination of plural forms, [...] when an article and noun, an adjective and noun, etc. follow each other, and (morphosyntactic) agreement should occur, only one item bears the plural marker [...]. one often hears *os homen* 'the(PL)[3] man(SG)' for *os homens* 'the(plural) men(plural)'; *as muyé* 'the (PL) woman(SG)' for *as mulheres* 'the(PL) women(PL)'; *duas boa pessoa* 'two(PL) good(SG) person (singular)' for *duas boas pessoas* 'two(PL) good(PL) persons(PL)'; *casas grande* '(lit,) houses(plural) big(SG)' for *casas grandes* 'houses(PL) big(PL)', etc.' (our translation).

3 Henceforth, the abbreviations PL and SG will be adopted to indicate plural and singular, respectively.

The morphological simplification evident in Popular Brazilian Portuguese is seen by Coelho as the result of massive contact between Portuguese and African languages, which highlights 'the extent of the phenomenon which, as with other phenomena in the Brazilian language, displays a creolizing tendency' (our translation) (1967, [1]1880–1886, 170). This relationship was also perceived by pioneering studies during the first decades of the 20th century. Mendonça (1933, 50ss.) spoke of the differentiation of the 'Brazilian dialect' from the Portuguese of Portugal, highlighting other differential factors 'which only intervened when the language was transplanted' to Brazil, resulting from the 'contribution of indigenous and African elements' (our translation). Mendonça (1933, 60) also stresses, as one of the most notable effects of this language contact, the absence of number agreement in the noun phrase and between the nominal predicate and its subject in copular predication.

However, the influx of three million migrants into Brazilian society between the late nineteenth and early 20th century, and the changes produced by the 1930 Revolution, which triggered the process of industrialization and urbanization, began to dilute the ethnic character of the linguistic divide of Brazil. Under the impact of Saussure's ideas, Brazilian philologists in the mid-20th century began to attribute secondary significance to the influence of language contact. In the work of Saussure (1916), the concept of language as a structure *où tout se tient*, with the history of a language being determined by the internal logic of the linguistic system, excluded the role of external factors such as language contact (Lucchesi 1998). Thus, on the one hand, Silva Neto (1988, [1]1957, 595) acknowledged the occurrence of processes of creolization of Portuguese during colonization, and attributed the loss of nominal agreement (*os livro* 'the (PL) book' (SG)', *as mesa* 'the (PL) table (SG)'), a fact "typical of creole languages", to the interference of language contact in "rural or dialectal language in Brazil". Silva Neto (1963, [1]1951, 166) stated that during the first century of colonization, a creole or semi-creole arose amongst Amerindians, Blacks and Mestizos ("índios, negros e mestiços"). Yet, he viewed this factor as secondary. Adhering to a particularly strict interpretation of the Saussurean vision, he concluded: 'the effect of speakers of other languages involves, in general, the acceleration of drift, that is, of tendencies already contained in the system' (1988, [1]1957, 604; our translation).

In a similar vein, the Brazilian structuralist Câmara Jr. (1975, 76–77) held that 'the black slaves have adapted themselves to the Portuguese language in the form of a creole variety' for 'on the large estates or farms of the colonial era and during the Empire, the contact of white masters with their black slaves was intense and close', which 'could explain certain innovations and simplifications of the Portuguese in Brazil in comparison with European Portuguese' (our translations). Yet, on the other hand, submitting to the structuralist framework, he concluded: 'It is clear, however, that deep phonological and grammatical changes would not occur without matching the structural trends of the Portuguese language itself' (loc. cit.; our translation).

In due course, the issue of language contact was soon resumed from another perspective: that of decreolization. Thus, the sociolinguist Guy (1981; 1989) argued that

the variation in contemporary vernacular Brazilian Portuguese resulted from an accelerated decreolization, beginning in the 18th century, of a Portuguese creole formed between the sixteenth and 17th centuries. The empirical basis for the proposal lay in the analysis of sociolinguistic variables which included the rules of nominal and subject-verb agreement, in a sociolinguistic corpus of the vernacular of illiterate speakers from Rio de Janeiro. His argument was underpinned by the notable sociohistorical parallels between Brazil and the creole-speaking regions of the Americas.

The creolist John Holm (1989; 1992), following up Guy's proposals, initially re-applied the term *semi-creole* to the vernacular varieties of BP, which would have evolved through

> 'contact between a non-creole (the colloquial and regional Portuguese brought from Europe to Brazil form de sixteenth century to the present) and a creole (the Portuguese based variety brought from São Tomé to Brazil during the sixteenth and seventeenth centuries, and latter varieties such as Helvécia Creole Portuguese' (Holm 1992, 62).

Subsequently, building on the views of Mello (1997), Holm (2004) proposed that vernacular Brazilian Portuguese was a language variety that had been partially restructured by the effects of an imperfect language shift involving Africans, Amerindians and their descendants.

The hypothesis of prior creolization has been opposed principally by Brazilian-based sociolinguists Naro/Scherre (1993; 2007), yet also by the creolists Parkvall/Álvarez López (2003). Naro/Scherre (1993; 2007) reject the idea of the prior existence of a creole in Brazil.[4] In their opinion, the primary source of variation in Brazilian Portuguese came from Portugal, originating in secular drift, and was carried further in Brazil as a result of language contact and shift:

> 'The Portuguese language spoken in Portugal before the colonization of Brazil already had a secular drift that drove it along a vector of development. [...] In Brazil, this vector met forces that reinforced it and expanded the original direction' (Naro/Scherre 2007, 47, our translation).

> 'the first origin of the phenomena of variable agreement rules came from Portugal, but the endemic conditions of pidginization and second language acquisition by adults, even before the arrival of (*African) slaves [...] accelerated and exaggerated the initial tendencies during the process of nativization by communities of the most diverse cultural baggage' (Naro/Scherre 2007, 52–53, our translation; *our insertion).

> 'The Portuguese language in Brazil did not need any structural influence from without in order to be transformed into Brazilian Portuguese' (Naro/Scherre 2007, 83, our translation).

4 For a brief discussion of their earlier formulation of this view, see Baxter (1998).

Thus, in this view the contact between languages would be a subsidiary factor. In this way, Naro/Scherre assume a theoretical framework partly resembling that adopted by Silva Neto, in the 1950s.

Since the notion of secular drift has been significantly challenged,[5] we will not discuss this matter here. The fundamental issue concerns whether input from Portugal was a necessary condition for the development of the variable agreement rules of Brazilian Portuguese. Naro/Scherre (1993; 2007) argue that the variable NP plural and subject-verb agreement of Popular Brazilian Portuguese have their origins, respectively, in a Romance phonetic tendency towards loss of word-final *-s* and *-m*. However, as Guy (1989) points out, it is unclear how such phonetic variation would become a widespread variation that affects the opposition between such distinct forms as *é*/*são* 'is/are' and *fez*/*fizeram* 'did (3SG)/did (3PL)'. Nor would such a phonetic origin account for why, in popular varieties, the person-number paradigm of the verb reduces to an invariant 1SG in contrast with a variably undifferentiated form matching 3SG, valid for all other person-number subjects (see section 5). The analyses of the structural configuration of the variation in nominal and verbal agreement in varieties of contemporary Brazilian Portuguese reveal that these phenomena are in nature morphosyntactic and not phonetic (Guy 1981; 1989; Lucchesi 2006a).

To demonstrate that variation in nominal and verbal agreement is not an exclusively Brazilian phenomenon, Naro/Scherre (2007, 49–70) point to the existence of instances of variation that extend from Old Portuguese to rural varieties of contemporary European Portuguese. However, the existence of a structured process of variation in Old Portuguese is unclear. With respect to current varieties of European Portuguese, recent variationist analyses by Gandra (2009) and Vieira/Bazenga (2013) demonstrate the peripheral character of the phenomenon.[6] As Mota et al. (2012, 166) observe: 'The absence of agreement markers in European Portuguese is quantitatively irrelevant and typologically limited – European Portuguese is characterized globally by the reiteration of the morphological marks of agreement'.

Thus, the variation in Portugal is best defined as a fluctuating phenomenon of linguistic performance, and not as a mechanism of variation in the structure of the grammar of the speech community, such as the broad structured variation observed in Popular Brazilian Portuguese. Indeed, Naro/Scherre (2007, 180) show that the general lack of subject-verb agreement occurring in Popular Brazilian Portuguese with all but 1SG, has no parallel in Portugal:

---

5 For example, Lightfoot (1979, 391) questions how drift determined by internal factors could proceed generationally: "Languages are learned and grammars constructed by the individuals of each generation. They do not have racial memories such that they know in some sense that their language has gradually been developing from, say, an SOV and towards a SVO type, and that it must continue along that path".

6 See also the review by Kabatek (2009).

> 'We did not find in the Portuguese documentation seen so far, any mention of the use of the verb ending of the 3rd person singular in the context of the 1st person plural in phrases such as *nós vai* 'we go(3SG)' [...]. Another case undocumented in Portugal is the variation *tu fala*/*tu falas* 'you (2SG) speak (3SG)/you (2SG) speak (2SG)', an item that can exist almost categorically without agreement in several Brazilian dialectal areas' (our translation).

Within the creolist camp, Parkvall/Álvarez López (2003) take issue with the earlier, and now superceded characterization of Popular Brazilian Portuguese as a *semi-creole*, proposed by Holm (1992). Firstly, they point out that certain creole-like structures in Brazilian Portuguese cannot be unambiguously related to creolization. Secondly, they argue that Holm's hypothesis regarding a historical connection between Brazilian Portuguese and São Tomé Creole Portuguese is not supported by historical documents.[7] Thirdly, drawing parallels with what is observed in the Caribbean, they suggest that the hypothesis of a rapid decreolization defended by Guy (1981) does not hold. In decreolization processes in the Caribbean, Rickford (1987) observes a continuum that stretches from typical local creole varieties to popular varieties of the target language.[8] In Brazil, Parkvall/Álvarez López (2003) assert that there is currently no creolized variety of Portuguese[9] and that no consistent records have come to light which clearly point to a linguistic variety of this kind in the history of Brazil, although scholars such as Silva Neto (1963, [1]1951) have made suggestive statements in this regard.

Given this situation, we embrace the views of Baxter (1998) and Parkvall/Álvarez López (2003) that it is not plausible to envisage a broad prior creolization process extant throughout Brazil, resulting in a single pan-Brazil creole which would have decreolized rapidly, resulting in the current varieties of popular Brazilian Portuguese. In this sense, we agree with the position of Rodrigues (2006, 155, note 3):

---

7 However, São Tomé was a staging post for the slave trade to Brazil, and slaves from São Tomé were certainly present in Bahia at different times. Thus, Antonil (2012, [1]1711) refers to their presence on sugar plantations in the Recôncavo region of Bahia, and they also appear in slave lists of the 19th century (Andrade 1988; Baxter/Lucchesi 1999).

8 Nevertheless, ongoing sociolinguistic research in Bahia, demonstrates that there do exist linguistic continua whereby certain morphology is being acquired, ranging from rural communities, through provincial towns and eventually to the urban center (Lucchesi 2015b).

9 The only report of a possible instance of creolization of Portuguese in Brazil was documented in 1961 by the dialectologist Carlota Ferreira, who interviewed two elderly direct descendants of the community of African slaves, at Helvécia, a village situated in the extreme south of Bahia (Ferreira 1988). The limited data collected at the time are inconclusive, although strongly point to a previous process of creolization (Baxter 1992; Megenney 1993). Ferreira did not return to the community, but the authors of this chapter recorded speech samples representative of the community in the 1980s and 1990s. The features in these samples do not constitute a creolized variety of Portuguese, but reflect broader structural changes, deeper than the ones found not only in Popular Brazilian Portuguese, but also in rural communities with similar characteristics, as discussed in section 4 of this chapter.

> 'It is not unlikely that some pidgins, or even Creole, have been developed in certain places, but without having achieved the stability that would allow them to expand into space and survive for a long time' (our translation).

Faced with this impasse, Lucchesi/Baxter/Ribeiro (2009) compiled a significant body of empirical research of the speech of rural communities formed by direct descendants of African slaves who still remain in relative isolation in rural Brazil. The research evaluates the influence of contact in the formation of localized varieties of popular Brazilian Portuguese, based on current theories of language contact and change. In the opinion of Faraco (2010, 294–296), the referred study:

> 'By emphasizing the effects of language contact and demonstrating the existence of only localized processes of intense irregular linguistic transmission (close to creolization), the work dissolves the traditional dichotomous discussion (prior creolization versus secular drift). At the same time, it makes significant contributions towards redirecting the focus of historical research of Brazilian Portuguese, shifting that focus away from the narrow path of analyzing the evolution of the learned varieties' (our translation).

This new perspective renewed the agenda of linguistic research on the issue. This agenda is further bolstered by input from current research on recently acquired and recently nativized varieties of Portuguese in multilingual situations in Africa, which also address variation.[10] On the one hand, the socio-historical research seeks to clarify the sociolinguistic context of the early centuries of the history of Brazil in which contact between languages affected the Portuguese language. On the other hand, the theoretical and empirical analysis seeks to identify which features of the popular Portuguese language result from changes induced by contact between languages. These two issues will be addressed in the following sections.

3 Language contact in the history of Brazil

From the outset, the Portuguese colonization of Brazil was based on forced labor, initially of the native indigenous population and then of enslaved and forcibly displaced Africans. These enslaved people constituted the majority of Brazil's population throughout the colonial period and the Empire. The sociolinguistic panorama of the first four centuries of history of Brazil constituted a diglossic situation. In this period

10 Unfortunately, much of the discussion of the origins of variation in Brazilian Portuguese has been conducted without reference to research on recently acquired and recently nativized varieties of Portuguese, such as those of São Tomé (Baxter 2004; Gonçalves/Hagemeijer 2015) and Mozambique (JonAnd 2011; Gonçalves 1996; 2010). Variation in agreement rules is common to each of these varieties, yet the target language is the standard of Portugal, with all its agreement rules complete, and the variation arises because of adult L2 acquisition in a contact situation.

only a third of the population, including the ruling elite, consisted of native speakers of Portuguese, themselves born of native speakers, whereas the other two thirds consisted of Africans, Amerindians and their direct and racially mixed descendants, who spoke Portuguese as a second language or as a nativized variety thereof, in a multilingual context.

In the early 16th century, the indigenous people of the coast of Brazil, from São Paulo to the mouth of the Amazon, spoke very closely related varieties of the Tupi-Guarani language family (Rodrigues 2006). These languages received several names derived from the Tupi radical, such as *tupinambá*, *tupinaé* and *tupiniquim* (or *tupinaquim*). Thus, the colonizers and the Jesuit missionaries referred to these varieties as *língua geral da costa do Brasil* (general language of the coast of Brazil), and the term *língua geral* prevailed. This *língua geral* was preserved in various parts of the coast of Brazil as a language of colonial intercourse and language of the Jesuit catechism (Freire/Rosa 2003).

In the region of São Paulo, Portuguese colonization began in 1532 with the founding of the village of São Vicente, on the coast. Subsequently, the settlers headed inland, founding the village that would become the present city of São Paulo. Initially, the colonization process aimed to enslave the indigenous population, and to this end conducted major expeditions called *bandeiras*, advancing into neighboring regions. On the sociolinguistic level, the small contingent of settlers, overwhelmingly men, produced an extensive process of miscegenation, forming a *mameluco*[11] society where use of the *língua geral* predominated. This diglossia between *língua geral* and the Portuguese of the colonists, in the São Paulo society of the 15th and 16th centuries, constitutes the first configuration of the linguistic divide of Brazil (Monteiro 1995, 165).

Another variety of *língua geral* came to dominate the colonial society that the Portuguese established in Maranhão, after the expulsion of French privateers, in 1615. The nucleus of this society, formed by Portuguese and Tupinambá Indians, expanded into the Amazon in search of jungle spices and enslaved further indigenous peoples, mostly speakers of different languages, including other language families, such as the Aruak and Karib. However, the language that prevailed in the colonization of the Amazon was the tupinambá *língua geral*, called *nheengatu* (lit. 'good language'). This language became nativized among many Amazonian indigenous peoples and it is still spoken in the Upper Amazon region, where it is an official language in the municipality of São Gabriel da Cachoeira, on the border with Colombia.

The discovery of gold and diamond deposits in the province of Minas Gerais, in the late 17th century changed the linguistic landscape of southeastern Brazil profoundly. Hundreds of thousands of Portuguese flocked to the regions of the states of Minas Gerais, São Paulo and Rio de Janeiro, during the 18th century, importing many hundreds of thousands of African slaves. This influx, logically reinforced the presence

11 The term *mameluco* refers to the mestizo children of European settlers with indigenous women.

of Portuguese and second language varieties thereof, yet also led to situations of African language maintenance (Byrd 2012; Petter 2006). As a result, the use of *língua geral* in this region declined substantially, almost disappearing in the 19th century.[12] Thus, the varieties of *língua geral* were progressively confined to remote and peripheral areas of Brazilian society. Needless to say, the Tupi languages, through the varieties of *língua geral*, left a huge mark on the referential lexicon of Brazilian Portuguese.

On the other hand, in the economically more dynamic regions of Portuguese America, such as the surroundings of the cities of Olinda and Salvador in northeastern Brazil, in the 16th and 17th centuries, the local indigenous population was quickly decimated, being soon replaced by large numbers of African slaves. Thus, the main driving force of the Portuguese colonial venture in Brazil was the workforce of African slaves and their descendants, referred to as *crioulos* (creoles).[13]

To ensure the conditions of domination, traders and slave owners sought to break all bonds of sociability the slaves possessed. Slaves of different ethnicities were mixed, from the moment of embarkation in Africa, to their arrival at the slave markets in Brazil, and on plantations; and the use of African languages was repressed to prevent uprisings (Mattoso 2003). The violent physical and symbolic oppression inherent in the slavery process was in good part responsible for the nativization of Portuguese by this large segment of Brazilian society, in the Colonial and Imperial periods. Thus, Africans and their descendants, assimilating the language of the colonizer, became agents for the diffusion of Portuguese in Brazil (Ribeiro 1995, 166), contributing decisively to the shaping of key aspects of Brazilian culture such as cuisine, religion, dance and music.

Not only in Brazil but in the entire process of colonization of America between the 16th and 19th centuries, the importation of a large African slave labor contingent played a crucial role. On the linguistic level, the African contribution is distinguished by the emergence of creole languages in the Caribbean region, in plantation societies formed on the basis of large agro-exportation properties employing slave labor. Among the more than thirty creole languages that developed in the region, we find Haitian, whose lexicon is of French origin, Jamaican, with an English-based lexicon, Papiamento, with a Spanish/Portuguese-based lexicon, and others, such as the English lexicalized Sranan and Saramaccan in Suriname. In the US Southeast, where cotton plantations based on African slavery were abundant, creole languages also developed, as exemplified by Gullah, whose descendent population is found today in the states of South Carolina and Georgia (Holm 1989, 491–494).

---

12 Contrary to the traditional view, we believe that the Pombaline linguistic policy, which forbade the use of *língua geral*, was only one factor that led to the decline of the *línguas gerais* in such a large country with severe limitations of administration in the rural areas. Socioeconomic-demographic changes, which were substantial and varied throughout the period in question, are significant factors.

13 Child of an African born in Brazil.

For over 350 years (1500–1866), the slave trade brought to the American continent about ten million Africans, nearly half of which landed in Brazil.[14] Already in the 17th century, approximately half of Brazil's population was composed of Africans and their descendants. This increased to 60 % in the 18th century and ranges between 57,5 % and 65 %, in the 19th century, as shown in the table below:

Table 1: Population of Brazil by ethnic group, 16th to 19th centuries (Mussa 1991, 163).

Ethnic group	1583–1600	1601–1700	1701–1800	1801–1850	1851–1890
Africans	20 %	30 %	20 %	12 %	2 %
Crioulos	–	20 %	21 %	19 %	13 %
Mulattoes	–	10 %	19 %	34 %	42 %
White Brazilians	30 %	5 %	10 %	17 %	24 %
Europeans	50 %	25 %	22 %	14 %	17 %
Integrated Amerindians[15]		10 %	8 %	4 %	2 %

Africans and their descendants formed the bulk of the workforce of sugarcane, tobacco and cotton plantations in the Brazilian northeast between the 17th and 19th centuries; the labor force for the extraction of gold and precious stones in Minas Gerais in the 18th century, and also that of the coffee plantations in the Paraíba Valley and the São Paulo plains, in the 19th century. The importation of slaves ceased officially in 1850, and slavery was abolished in Brazil only in 1888. Until the beginning of the 20th century, the vast majority of this massive population of African descent resided in the rural sector and was illiterate.

This very composition of Brazilian society should apparently favor the emergence of a broad creolization process of Portuguese, but none of the available evidence supports this. A careful observation of the socioeconomic and demographic contexts of Brazil's history clarifies the issue. There are significant differences between the formation of Brazilian society and the formation of the Caribbean plantation societies, beginning with the size and complexity of the former in comparison with the latter, which may explain why an extensive creolization of Portuguese did not occur in Brazil.

In Brazil, the proportion of speakers of the dominant language, Portuguese, was always more than 30 %, as can be seen in Table 1 above. Therefore, the proportion of the dominant group in Brazil was always higher than the total of 20 % that Bickerton (1981) proposed as the maximum for creolization to occur. Between the 16th and 18th centuries, in Brazil, just as in the Caribbean, large agro-export enterprises constituted the driving force of colonization, employing African slave labor and exporting highly

14 Source: The Trans-Atlantic Slave Trade DataBase, http://slavevoyages.org/tast/index.faces, last accessed 14.02.2022.

15 Amerindians no longer living in tribal conditions.

valued products to Europe – sugar, tobacco and cotton – boosting the European economy. However, the strongly polarized master-slave relationship typical of plantation societies, is mediated in Brazil, by a large contingent of cane growers and food producers (mainly cassava flour), comprised of minor land owners and farmers with very small number of slaves (one to five). As Ferlini (2002, 25–26) points out, in the immediate area around a large sugarcane plantation, which could have more than 200 slaves, there resided a large number of small-scale farmers, in an interdependent relationship with the large plantation.

Extensive livestock farming was also integrated into this agro-export system, in food production, and it too employed a small number of African slaves (Linhares 2002, 109–122). This more complex and more nuanced structure than that of a prototypical plantation society would explain why there was no widespread creolization of Portuguese, although localized processes of creolization may have occurred, particularly in the context of plantations that had more autonomy, and in major *quilombos* 'maroon communities' such as Palmares (Lucchesi 2009). However, these manifestations of pidginization/creolization did not last.

Another factor that inhibited creolization in Brazil was the mining boom that occurred in the 18th century. Between 1700–1800, Brazil's population grew 11 times over, greatly increasing the proportion of native Portuguese speakers. In addition, the wealth of gold and diamonds produced the first major instance of urbanization of Brazilian society, with its agencies of standardization and linguistic diffusion (Lucchesi 2006b). Finally, unlike the plantation context, the situation of the slave in mining did not favor creolization (Castro 1990, 108). Proof of this is that in regions where large masses of African slaves were concentrated in mining in South America, such as Peru, Bolivia and Colombia, there were no creolization processes, in contrast to the Caribbean, where slaves were employed on plantations.

However, it is widespread miscegenation, a principal characteristic of Brazilian society, which is the main reason why an extensive process of creolization did not occur. This is due to the *sui generis* characteristic of the patriarchal society that Portuguese colonization produced in Brazil, outlined in the classical analysis of Gilberto Freyre (1936). The widespread miscegenation of masters and slaves reached huge proportions in Brazil, as shown in Table 1. The percentage of mulattos (children of white fathers and African or *crioulo* mothers) already constituted more than a third of population in the 18th century and it became more than 40 % of the population in the 19th century. Despite the violence of slavery, this promiscuous relationship between masters and slaves favored greater integration, especially of those of mixed race African descent, but also of the *crioulos* in general, reducing the segregation which is the basis of the creolization processes, as discussed in the next section.

The fact that Portuguese did not creolize extensively in Brazil does not mean that contact between languages was insignificant. The current popular varieties of Brazilian Portuguese have their origin in the precarious acquisition of Portuguese as a second language by millions of acculturated Amerindians and enslaved Africans and the

subsequent nativization of this divergent model of second language among their descendants. This process of nativization has been termed *irregular language transmission* (Baxter/Lucchesi 1997; Lucchesi 2008; Lucchesi/Baxter 2009).[16] And although its results in Brazil are not so deep as to constitute creolization, they were sufficient to produce significant changes in the grammar of the linguistic variety thus formed, particularly in relation to certain syntactic mechanisms, such as rules of agreement, which express redundant information, or mechanisms devoid of informational value, such as the reflexive pronoun *se* in such verbs as *sentar-se* 'to sit'.[17]

4 Irregular linguistic transmission

The concept of irregular linguistic transmission is based on a broad vision of the induction of grammatical changes in situations of massive contact between languages as a gradual and variable process. On the one hand, such contact situations can generate a historically new language, a pidgin or creole, which has a grammar qualitatively different from that of the target language (Rougé 2008). On the other hand, the result may be merely the development of a new historical variety of the dominant language, which has structural characteristics similar to those of creole languages. In the case of pidgins and creole languages, irregular linguistic transmission is radical, whereas in the second case, the process is more superficial, in the sense that the grammar of the emergent language is less re-structured.[18] However, in both instances, what is essentially at stake is the need to build the grammatical structures that were not incorporated during the precarious process of acquiring the language of the dominant group by adult speakers of the subjugated population.

As such, the key factor determining the outcomes of irregular linguistic transmission is the degree and type of input from the subjugating language in the initial phase of contact, conditioned by restricted access to L1 models and ready access to the L2 contact models of adult colleagues. For the occurrence of the radical restructuring characteristic of pidginization and creolization, access to the target language models must continue to be restricted during the formation of the new speech community. This is determined by very specific sociolinguistic contexts, such as those found in certain plantation societies, or in maroon communities. In the Americas, these situations can be summarized as follows:

16 Similar hypotheses regarding this type of nativization in contact situations have been presented by Winford (2003) and Holm (2004).

17 The loss of semantically empty reflexive pronouns is also reported in the Portuguese of Mozambique, which is mainly spoken as a second language (Gonçalves 1996).

18 Holm (2004) presents a somewhat similar proposal, by means of his concept of partial linguistic restructuring.

(i) removal of populations from their original cultural and linguistic context, as occurred with the African slave trade;
(ii) concentration of a large linguistically heterogeneous contingent under the control of a dominant, numerically inferior group;
(iii) segregation of the new community formed in the contact situation.

The numerical superiority of the subjugated group and segregation greatly restrict access to the models of the dominant language, setting the stage for the introduction of new grammatical elements. Such items may originate from the grammars of the substrate languages (Lefèbvre 1998; Lumsden 1999; Siegel 2008), or they may result from grammaticalization processes guided by universal psycholinguistic factors (Bickerton 1999; Adone 2012). Therefore, the degree of grammatical restructuring occurring in irregular linguistic transmission processes is proportional to socio-historical and demographic factors, as the originality of the grammatical structure of pidgins and creole languages stems from a restrictive social context that disfavors and impedes acquisition of the grammar of the subjugating language.

In a different socio-historical context, greater access of the speakers of other languages and their descendants to the grammatical models of the target language inhibits substrate transfer processes and grammaticalization, crucial for the formation of pidgins and creoles. Therefore, the greatest difference between the processes of pidginization and creolization and lighter irregular linguistic transmission processes is that in the former case, the grammar of the language being developed in the contact situation is basically formed by exogenous elements, whereas in the second case, the grammatical elements of the language of the dominant group outweigh any embryonic processes of grammaticalization and substrate transfer.

However, it must be kept in mind that, in both cases, the development of varieties of second language among adult speakers of the dominated groups entails incomplete acquisition (erosion) of morphology of the target language, a fact amply attested in the literature on language contact (Labov 2007; Siegel 2006; 2008; Trudgill 2010; Winford 2003). Even when this erosion is lighter, it still compromises the items most commonly affected in massive language contact: verbal person and number inflection, gender and number agreement in the noun phrase, and pronominal case (Koefoed/van Marle 2004; Trudgill 2009). With respect to these items, a further distinction between creolization and lighter irregular linguistic transmission is evident. In the creolization, these morphemes are essentially eliminated (resurfacing during decreolization); yet in lighter irregular language transmission, rather than being completely eliminated, these same morphemes are retained and subject to an extensive process of variation.

Formulated in these terms, the concept of irregular linguistic transmission can account for an ample spectrum of situations and products of language contact in the history of the colonization of the Americas. The transplantation of European languages to the American continent entailed many language contact situations which led to the

retention of features common to known creole languages. Since the radical effects of contact on the structure of European languages are clear in situations of creolization, it is unreasonable to postulate that similar contact situations did not affect the formation of the popular varieties of English, Spanish or Portuguese in the Americas, and that this effect was simply restricted to a mere acceleration of evolutionary trends already present in the secular drift of these languages.[19] Thus, varieties of English, such as African-American Vernacular English, and Spanish, in various parts of the American continent, currently present linguistic traits of a creole type, such as the absence of copula, simplification of inflectional morphology, and even embryonic processes of grammaticalization of particles of tense-mood-aspect (Clements 2009; Holm 2004; Lipski 2008), without constituting a creole language as such. Similarly, popular varieties of Brazilian Portuguese also exhibit characteristics which clearly point to processes of change produced by contact between languages during their development.

5 Characteristics of Popular Brazilian Portuguese derived through language contact

In Brazil, there is a clear distinction between the speech of the literate elite and that of the poor living on the outskirts of large cities and in the rural interior. The latter variety suffers strong prejudice, which is reinforced by the mass media, emphasizing the linguistic division of the country. This sociolinguistic polarization reflects the socioeconomic structure of Brazilian society, marked by a radically imbalanced distribution of wealth, workforce exploitation and marginalization of the poor. In this context, linguistic discrimination constitutes a powerful instrument of domination of the ruling classes (Lucchesi 2015a).

The features of the popular language that suffer discrimination are those that resulted from changes triggered by language contact (see previous section), and which reveal the racist character of linguistic discrimination in Brazil (Lucchesi 2011). Such features are present both at the phonological and syntactic levels. Of the latter, the most notable is the reduction of subject-verb and nominal agreement in popular speech, while the literate elite apply these rules at a much higher frequency. The differences are indicated in examples (1) and (2). In (1), in Standard Portuguese, number agreement in the NP and Subject-verb agreement are applied:

19 In diverse locations in Latin America contact with indigenous languages or African languages has induced contact-induced change/variation. Where there were plantation societies, more radical change occurred (English/French/Spanish-based creoles).

(1) Meu**s** filho**s** trabalha**m** muito.
1SG.POSS-M-PL child-M-PL work-PRS-3PL much
(Standard Portuguese)
'My sons work a lot'

However, example (2), from Popular Brazilian Portuguese, shows number in the NP marked only on the first item of the NP, and the verb does not agree in number and person with the subject NP:

(2) Meu**s** filho trabalha muito.
1SG.POSS-M-PL child-M-SG work-PRS-3SG much
'(lit.) My sons works a lot'

Where subject-verb agreement is concerned, urban university-educated Brazilians also employ the rule variably, but at virtually categorical levels (Araújo 2014). However, the frequency of rule application decreases as one passes from the higher social classes to the lower social classes with less education. The lowest levels of rule application are found in isolated rural communities formed by the direct descendants of African slaves (Lucchesi 2004, many of which have their origins in maroon communities. They are poor communities depending on subsistence agriculture, and their members have little or no education. Their speech is a specific variety of Brazilian Portuguese, having been more affected by contact between languages and, for this reason, it is referred to as Afro-Brazilian Portuguese (Lucchesi/Baxter/Ribeiro 2009).[20]

This sociolinguistic distribution supports the hypothesis that the loss of person and number inflection on the verb is a consequence of the morphological simplification produced by language contact. The greater the historical relationship of the speech community to language contact, the lesser the use of the agreement rule. The difference is not only quantitative, it is qualitative as well. In popular Brazilian Portuguese, the change in the paradigm affects all person forms except the 1st person singular. In Afro-Brazilian speech communities, for which there is evidence of changes closer in level to actual creolization, such as in the Helvécia community in Bahia State,[21] the variation in the use of the agreement rule affects all person-numbers in the paradigm, as shown below:

20 Lipski (2008) also uses the term *Afro-Bolivian Spanish* for Spanish-based linguistic varieties with histories of development and formal characteristics of a similar nature to those discussed here.
21 See footnote 10.

Table 2: Person and number in three varieties of Brazilian Portuguese.

English	Standard Brazilian Portuguese	Popular Brazilian Portuguese	Afro-Brazilian Portuguese
I work	eu trabalh**o**	eu trabalh**o**	eu trabalh**o**(a)[22]
you work	você trabalha	tu/você trabalha	tu trabalha
he/she works	ele/ela trabalha	ele/ela trabalha	ele/ela trabalha
we work	nós trabalha**mos**	nós trabalha(**mo**(**s**))	nós trabalha(**mo**(**s**))
you work	vocês trabalha**m**	vocês trabalha(**m**)	vocês trabalha(**m**)
they work	eles/elas trabalha**m**	eles/elas trabalha(**m**)	eles/elas trabalha(**m**)

Further evidence in favor of the language contact hypothesis is found in the parallel with creolization processes of Portuguese in Africa, wherein the verbal inflection of Portuguese was not acquired. This is seen in the comparison of popular Portuguese and Afro-Brazilian Portuguese with Forro, the Portuguese Creole spoken in São Tomé (Africa) (Ferraz 1979; Hagemeijer 2013):

Table 3: The verb 'know' in Forro, Afro-Brazilian Portuguese and Popular Brazilian Portuguese.

English	Forro	Afro-Brazilian Portuguese	Popular Brazilian Portuguese
I know	n'sebê	eu sei/sabe	eu sei
you know	bo sebê	tu/você sabe	tu/você sabe
he/she know**s**	e sebê	ele/ela sabe	ele/ela sabe
we know	nõ sebê	nós sabe(**mo**(**s**))	sabe(**mo**(**s**))
you know	nãsse sebê	vocês sabe(**m**)	vocês sabe(**m**)
they know	inem sebê	eles/elas sabe(**m**)	eles/elas sabe(**m**)

The above table demonstrates what was predicted in the previous section. In the case of Forro, a creole formed by a radical process of irregular linguistic transmission, verbal person-inflection has simply not been acquired. In the case of lighter irregular linguistic transmission, substantial variation is observed in the use of such morphology, yet the morphology is not is partially acquired. Furthermore, the change is deeper and

22 The brackets indicate the segment affected by the variation process.

more extensive in the varieties most affected by language contact, such as the Afro-Brazilian Portuguese in comparison with more common varieties of Popular Brazilian Portuguese.

A similar picture is found regarding noun phrase number agreement, where the plural in Popular Brazilian Portuguese varieties is preferentially marked only on the determiner, other items being preferentially unmarked: *uma***s** *coisa velha* (Popular Brazilian Portuguese) versus *uma***s** *coisa***s** *velha***s** (Standard Brazilian Portuguese).[23] Just as Coelho (1967, [1]1880–1886) had observed in the late 19th century (cf. section 1), this is another phenomenon that associates varieties of Popular Brazilian Portuguese with language contact. In isolated Afro-Brazilian rural communities, the lack of NP number agreement occurs in more than ninety percent of cases, whereas educated speakers in large urban centers apply the rule with a frequency greater than 70 % (Lucchesi 2013). Once again the parallel with the Portuguese Creoles of Africa is significant, as can be seen in example (3)a, from Cape Verde Creole, in which the plural is marked only in the determiner, as in Popular Brazilian Portuguese, in example (3)b, unlike the agreement mechanism in Standard Brazilian Portuguese, in example (3)c:

(3) a. ke**s** mininu tá fla tcheo.
DET-**PL** child ASP speak much
'Those children (lit. child) talk a lot'

b. Aquele**s** menino fala muito.
DET-M-**PL** child-M-**SG** speak-PRS-3**SG** much
'(lit.) Those child talks a lot'

c. Aquele**s** menino**s** falam muito.
DET-M-**PL** child-M-**PL** speak-PRS-3**PL** much
'Those children talk a lot'

23 In studies of noun phrase plural marking based on urban corpora, two proposals have been advanced: Guy (1981), working with Rio de Janeiro data from low literacy speakers of humble socioeconomic background, drew attention to the connection between linear position and grammatical class. However, Scherre (1988), also working with urban Rio de Janeiro data, but from a wider social class spectrum, reported that plural marking is dependent on linear position, with a tendency to mark the first item of the NP, and not dependent on grammatical class (see discussion in Baxter 2009). However, Andrade (2003) and Baxter (2009), working with rural Afro-Brazilian data, found that the majority of NPS in their corpus consisted only of two items: DET+N, where DET is represented by articles, demonstratives and possessives. In both studies, the predominant locus of plural marking (94 %) was the DET. It may be concluded from this that the nature of plural marking depends very much on the projection possibilities of the NP in the variety of Portuguese in question. In Popular Brazilian Portuguese varieties, and notably in the rural area, the association of PL with DET is much more apparent than in urban data that includes a wider spectrum of socioeconomic classes

The situation regarding NP number agreement parallels that of subject-verb agreement. Whereas in the creole the agreement rule was completely eliminated, in Popular Brazilian Portuguese a broad process of variation is found, which is most intense in those varieties most affected by language contact, such as Afro-Brazilian Portuguese. The difference between Afro-Brazilian Portuguese and Popular Brazilian Portuguese in general is particularly clear in the case of variation in gender agreement within the NP. This phenomenon does not occur in Popular Brazilian Portuguese, but does occur in Afro-Brazilian Portuguese, in communities most directly affected by language contact, such as the Helvécia community, where sentences such as (4) may be observed, where the masculine form of the indefinite article occurs with a feminine gender noun (Lucchesi 2003):

(4) Ás vezes duece um pessoa.
to+DEF-F-PL time.F-PL ail-PRS-3SG INDF-**M**-SG person.**F**-SG
'sometimes a person becomes ill'

Once again, the parallel with the Creole Portuguese languages is notable, as can be seen in Kriol, spoken in Guinea-Bissau:

(5) I tene **un** *fiju* *femya* **bonitu**.
3NOM-SG have-PRS-3SG INDF child female pretty
'(s)he has a pretty daughter (lit. a pretty female child)'

In this example, while *un*, *fiju* and *bonitu* derive from words inflected for masculine gender in Portuguese, in spite of the presence of *femya*, derived from a feminine gender word in Portuguese, there is no agreement rule assigning feminine gender to the former items.

Another process of morphological simplification that language contact produced in Popular Brazilian Portuguese is the variation in pronominal case inflection:

(6) a. Ele viu **eu**.
3NOM-M-SG see-PST.PFV-3SG 1NOM (=ACC)-SG
'he saw me (lit. he saw I)'

b. Ele **me** viu.
3NOM-M-SG 1ACC-SG see-PST.PFV-3SG
'he saw me'

(7) a. Maria procurou **tu**
Maria seek-PST.PFV-3SG 2SG-NOM (=ACC)
na feira.
LOC+DEF-F-SG market-F-SG
'Maria looked for you at the market'

b. Maria **te** procurou na
Maria 2SG-ACC seek-PST.PFV-3SG LOC+DEF-F-SG

feira.
market-F-SG
'Maria looked for you at the market'

Examples (6)a e (7)a, with pronouns *eu* and *tu* based on the nominative case form representing the accusative case, are not found in the speech of the literate elite, as in (6)b and (7)b, which only displays variation with the 3rd person, as in (8)a, in contrast with (8)b, in Standard Portuguese:

(8) a. João avisou **ela**
João warn-PST.PFV-3SG 3NOM(=ACC)-F-SG
do problema.
of+DEF-M-SG problem-M-SG
'João warned her (lit. she) of the problem'

b. João **a** avisou do
João ACC-F-3SG warn-PST.PFV-3SG of+DEF-M-SG
problema.
problem-M-SG
'João warned her of the problem'

This occurs because of the inflectional characteristics of this pronoun, which make it similar to a noun,[24] and the weak phonetic salience of the third person accusative clitic, which consists only of a vowel, unlike the other object clitics, which consist of a standard CV(C) syllable: *me* 1SG direct or indirect object, *te* 2SG direct or indirect object, *se* 3SG REFL, *lhe* 3SG indirect object, *nos* 1PL direct or indirect object, *vos* 2PL direct or indirect object (Câmara Jr. 1972, 47–53).

In Popular Brazilian Portuguese, the lack of case inflection affects all pronouns (Lucchesi/Mendes 2009). Once again, the parallel with the Creole Portuguese languages of Africa is remarkable, as can be seen in the examples of Cape Verde Creole, in (9) and (10). Here, case inflection is absent in the personal pronouns, which display the same form of the pronoun in subject and complement functions, with a reduction in phonetic form of the pronoun in object position when cliticized to the verb:

24 As with nouns, the third person pronoun inflects for gender and number: *ele*/*ele**s*** 'he/they (MASC)', *el**a***/*el**as*** 'she/they (FEM)'.

(9) a. **Mi** e fliz.
1SG COP happy
'I am happy'

b. El dà-**m** un livr.
3SG give.PST-**1SG** one book
'He gave **me** a book'

(10) a. **Bo** ta faze izarsísi.
2SG ASP do exercise
'You do the exercise'.

b. El dà-**b'** un livr.
3SG give.PST-**2SG** one book
'He gave you a book'.

In Brazilian Portuguese, variation in verbal and nominal number agreement occurs both in the speech of the educated elite, and in popular speech, but there is a huge difference between these two groups in terms of the frequency of the variation. In the popular variety, forms without agreement predominate, whereas in the speech of the elite, lack of agreement is highly restricted. Variation in gender agreement is found only in the Afro-Portuguese variety and variation in first and second person pronominal case inflection is found widely in rural varieties of Popular Brazilian Portuguese. All of these forms of popular speech are evaluated negatively by educated middle and upper-class speakers. These differences in use and social assessment, and the different trends of ongoing change in each social group (as discussed below), constitute the sociolinguistic polarization of Brazil (Lucchesi 2015a).

Yet other phenomena of popular Brazilian Portuguese can be related historically to contact between languages, in variation processes that affect diverse areas of the grammar: relative clauses, the shape of interrogative words, number and gender agreement in copular predicates and passives, bare nouns, reflexive pronoun use and forms of indeterminacy of the subject (Lucchesi/Baxter/Ribeiro 2009). In view of this, it can be concluded that, in cases of light irregular linguistic transmission, such as that which typifies the development of popular Brazilian Portuguese, the contact-induced changes are characterized by morphological simplification without, however, resulting in the elimination of grammatical devices as in the formation of the African varieties of Creole Portuguese,[25] but producing a broad and deep process of variation.

25 To mention just three examples, none of the Portuguese-lexified creoles of Africa, retained the Portuguese mechanisms of tense-mood-aspect inflection, NP plural agreement or subject-verb agreement. In the creoles, tense-mood-aspect expression is indicated principally by independent particles.

Grammaticalization, and transfer of grammatical structures of the substrate languages, typical of creolization are infrequent in the lighter type of irregular linguistic transmission that produced popular Brazilian Portuguese. The result is the development of a variety of the dominant language, and not a new language with a qualitatively distinct grammar,[26] as in creolization. Nevertheless, marginal developments related to substrate transfer and original grammar restructuring, more typical of creolization, can be found in popular varieties of Brazilian Portuguese.

One such case is that of *dative alternation*. In the grammar of isolated Afro-Brazilian rural communities, and in Popular Brazilian Portuguese, in addition to the prepositional dative construction (PDC) of Standard Portuguese, exemplified in (11), there is an alternative: the double object construction (DOC), exemplified in (12):

(11) Eu dei *o* *remédio*
1SG-NOM give-PST.PFV-3SG DEF-M-SG medicine-M-SG
aos **meninos**.
to+DEF-M-PL child-M-PL
'I gave the medicine to the children'

(12) Eu dei **os** **menino**
1SG-NOM give-PST.PFV-3SG DEF-M-PL child-M-PL
o *remédio*.
DEF-M-SG medicine-M-SG
'I gave the children the medicine'

The DOC, which is present in Germanic languages such as English and Dutch, is not part of the grammatical repertoire of the Romance languages, so that, for an urban Brazilian speaker, the construction exemplified in (12) is ungrammatical.[27] However, DOCs are typical of most of the African languages once present in Brazil, including Ewe-Fon and most Bantu languages (Baxter/Mello/Santana 2014). Thus, the DOC's presence in popular varieties of Brazilian Portuguese can be seen as a case of original grammar restructuring triggered by linguistic contact (Lucchesi/Mello 2009). This hypothesis is strengthened by the fact that the DOC is widespread among the creole languages of the Atlantic region (Parkvall 2000), even in those derived from Romance languages, as in example (13), from Fa d'Ambu, the Portuguese Creole of Anobon island, in the Gulf of Guinea:

26 The grammatical differences mentioned in the previous footnote are examples of this qualitative distinction.

27 This observation is based on: (i) the rarity of such structures in urban sociolinguistic corpora of Salvador; (ii) acceptability judgements of our (mainly urban Afro-Brazilian) undergraduate and postgraduate students of syntax at the Federal University of Bahia; and, (iii) consultation with specialists in syntax and sociolinguistics from other regions in Brazil.

(13) Malía da pe-d'eli tabaku.
Maria give.PST father.GEN.3SG tobacco
'Maria gave her father tobacco'

On the other hand, sociolinguistic analyses of these variable phenomena in Popular Brazilian Portuguese have revealed a complex scenario of language change in progress, in Labovian terms (Labov 1994), wherein the form that originated in language contact is being replaced by the standard form (Lucchesi 2015b). In the specific case of Afro-Brazilian rural communities, younger members, who have a higher level of education and are better placed in the labor and consumer markets than older speakers, lead the process of change (Lucchesi/Baxter/Ribeiro 2009). This scenario can be extended to all the popular classes, in a linguistic leveling process in which the linguistic forms of prestige in large urban centers are spread to all regions of the country, through the mass media, education and demographic movement. This tendency of popular Brazilian Portuguese to replace older, contact-induced forms with urban standard forms, contrasts with the tableau of stable variation observed in the middle and upper classes, in cases such as the agreement rules (Araújo 2014). This difference in the tendencies of linguistic change also constitutes one of the parameters for defining the sociolinguistic polarization of Brazil (Lucchesi 2015a).

6 Conclusion

Brazil today is a linguistically divided country, and it has been so ever since the beginning of its colonization by the Portuguese. In the early stages of colonization, a diglossic situation existed between the colonial elite and the mass of enslaved Indians and Africans. Yet, throughout the colonial period and the Empire, the Portuguese language was imposed and acquired as a second language by Indians and Africans to gradually become the first language of their descendants. This process of nativization through irregular linguistic transmission produced profoundly altered varieties of Portuguese, yet without yielding the more radical effects of creolization, except in isolated, ephemeral cases, which did not last. With the assimilation or extermination of the greater part of the indigenous population and the official end of the slave trade in 1850, the sources of multilingualism in Brazilian society all but vanished. Its place was filled by a sociolinguistic divide with the language of the Brazilian elite, heavily influenced by the language of the former colonial power, in opposition to nativized varieties of Portuguese of descendants of Africans and Amerindians, strongly altered by language contact (Lucchesi 2009; 2015a).

However, widespread miscegenation and the arrival of some three million European and Asian immigrants between the second half of the 19th century and the first half of the 20th century helped dilute the ethnic character of the sociolinguistic divide (Lucchesi 2001). Moreover, this sociolinguistic cleavage began to be mitigated by the

thorough process of industrialization and urbanization of Brazilian society, commencing effectively with the Revolution of 1930. This led to the migration of vast rural masses to the cities. Today more than eighty percent of the population lives in urban centers, or in their periphery. Urbanization, education and the mass media promote the spread of Standard Brazilian Portuguese, in a linguistic leveling process that gradually tends to eliminate characteristics of Popular Brazilian Portuguese produced by past language contact.

Nevertheless, since radical inequality in the distribution of wealth in Brazil and marginalization of the poor severely constrain this linguistic leveling, the linguistic markers of former language contact still exist in popular speech, and are subject to linguistic discrimination by the middle and upper classes. This reinforces the sociolinguistic polarization, the origins of which date back to the dawn of Brazilian society (Lucchesi 2015a). However, sociolinguistic analyses reveal linguistic change in progress in the lower classes, whose younger members, who now have more education and are better placed in the labor and consumer markets, use urban standard forms more than do the older speakers of the same social class, whose speech retains markers of past language contact. It is plausible that, with economic and social progress in Brazil, the traces of past contact between Portuguese, Amerindian and African languages will in some cases disappear, and in other cases be retained as an integral part of the Brazilian Portuguese system.

7 References

Adone, Dany (2012), *The acquisition of creole languages: how children surpass their input*, Cambridge, Cambridge University Press.

Andrade, Maria José de Souza (1988), *A mão de obra escrava em Salvador, 1811–1860*, São Paulo, Corrupio.

Andrade, Patrícia Ribeiro de (2003), *Um fragmento da constituição sócio-histórica do português do Brasil: variação na concordância nominal de número em um dialeto afro-brasileiro*, Master's thesis, Salvador, Universidade Federal da Bahia.

Antonil, André João (2012, [1]1711), *Cultura e opulência do Brasil*, Salvador, Secretaria de Cultura do Estado da Bahia.

Araújo, Silvana Silva de Farias (2014), *A concordância verbal no português falado em Feira de Santana-Ba: sociolinguística e sócio-história do português brasileiro*, Doctoral thesis, Salvador, Universidade Federal da Bahia,.

Baxter, Alan (1992), *A contribuição das comunidades afro-brasileiras isoladas para o debate sobre a crioulização prévia: um exemplo do estado da Bahia*, in: Ernesto d'Andrade/Alain Kihm (edd.), *Actas do Colóquio Internacional sobre Línguas Crioulas de Base Portuguesa*, Lisboa, Colibri, 7–35.

Baxter, Alan (1998), *O português vernáculo do Brasil—Morfossintaxe*, in: Matthias Perl/Armin Schwegler (edd.), *América negra: panorámica actual de los estudios lingüísticos sobre variedades hispanas, portuguesas y criollas*, Frankfurt am Main/Madrid, Vervuert/Iberoamericana, 97–134.

Baxter, Alan (2009), *A concordância nominal*, in: Lucchesi, Dante/Baxter, Alan/Ribeiro, Ilza (edd.), *O Português Afro-Brasileiro*, Salvador, EDUFBA, 269–293.
Baxter, Alan (2004), *The development of variable NP plural agreement in a restructured African variety of Portuguese*, in: Geneviève Escure/Armin Schwegler (edd.), *Creoles, contact and language change. Linguistics and social implications*, Amsterdam/Philadelphia, Benjamins, 97–126.
Baxter, Alan/Lucchesi, Dante (1997), *A relevância dos processos de pidginização e crioliza̧ção na formação da língua portuguesa no Brasil*, Estudos Lingüísticos e Literários 19, 65–84.
Baxter, Alan/Lucchesi, Dante (1999), *Un paso más hacia la definición del pasado criollo del dialecto afro-brasileño de Helvecia (Bahia)*, in: Klaus Zimmermann (ed.), *Actas del congreso internacional sobre lenguas criollas de Base Española y Portuguesa*, Berlin, Instituto Ibero-Americano, 119–141.
Baxter, Alan/Mello, Camila/Santana, Natali (2014), *A construção de objeto duplo e as influências do substrato no português afrobrasileiro (e africano)*, Papia 24/2, 283–306.
Bickerton, Derek (1981), *Roots of language*, Ann Arbor, Karoma.
Bickerton, Derek (1999), *How to acquire language without positive evidence. What acquisitionists can learn from creoles?*, in: Michel DeGraff (ed.), *Language creation and language change: creolization, diachrony, and development*, Cambridge, The MIT Press, 49–74.
Byrd, Steven (2012), *Calunga and the Legacy of an African Language in Brazil*, Albuquerque, University of New Mexico Press.
Câmara Jr., Joaquim Mattoso (1972), *Dispersos*, Rio de Janeiro, Fundação Getúlio Vargas.
Câmara Jr., Joaquim Mattoso (1975), *História e estrutura da língua portuguesa*, Rio de Janeiro, Padrão.
Castro, Yeda Pessoa de (1990), *Os falares africanos na interação social dos primeiros séculos*, in: Linalda Arruda Mello (ed.), *Sociedade, Cultura & Língua*, João Pessoa, Shorin.
Clements, J. Clancy (2009), *The Linguistic Legacy of Spanish and Portuguese—Colonial Expansion and Language Change, Cambridge*, Cambridge University Press.
Coelho, Adolfo (1967, [1]1880–1886), *Os dialectos românicos ou neo-latinos na África, Ásia e América*, Boletim da Sociedade de Geografia de Lisboa, 1–234.
Faraco, Carlos Alberto (2010), *Recensións*, Estudos de Lingüística Galega 2, 292–296.
Ferlini, Vera Lúcia ([2]2002), *Pobres do açúcar. Estruturas produtivas e relações de poder no Nordeste colonial*, in: Tamás Szmrecsányi (ed.), *História Econômica do Período Colonial*, São Paulo, Hucitec, 21–34.
Ferraz, Luiz Ivens (1979), *The Creole of São Tomé*, Johannesburg, Witwatersrand University Press.
Ferreira, Carlota (1988), *Remanescentes de um falar crioulo brasileiro*, in: Carlota Ferreira et al. (edd.), *Diversidade do Português do Brasil*, Salvador, UFBA, 21–32.
Freire, José Ribamar Bessa/Rosa, Maria Carlota (2003), *Línguas Gerais: política linguística e catequese na América do Sul no período colonial*, Rio de Janeiro, EDUERJ.
Freyre, Gilberto (1936), *Casa grande e senzala*, Rio de Janeiro, Schmidt.
Gandra, Ana Sartori (2009), *A concordância verbal no português europeu rural*, in: Klebson Oliveira/Hirão Cunha e Souza/Luís Gomes (edd.), *Novos Tons de Rosa*, Salvador, EDUFBA, 142–161.
Gonçalves, Perpétua (1996), *Aspectos da sintaxe do português de Moçambique*, in: Isabel Hub Faria et al. (edd.), *Introdução à linguística geral e portuguesa*, Lisboa, Caminho, 313–322.
Gonçalves, Perpétua (2010), *A génese do português de Moçambique*, Lisboa, Imprensa Nacional—Casa da Moeda.
Gonçalves, Rita/Hagemeijer, Tjerk (2015), *O português num contexto multilingue: o caso de São Tomé e Príncipe*, Revista Científica da UEM 1/1, 87–107.
Guy, Gregory (1981), *Linguistic Variation in Brazilian Portuguese. Aspects of Phonology, Syntax and Language History*, Doctoral thesis, Philadelphia (PA), University of Pennsylvania.
Guy, Gregory (1989), *On the nature and origins of Popular Brazilian Portuguese*, in: *Estudos sobre el Español de América y linguística afroamericana. Ponencias presentadas en el 45 Congreso*

Internacional de Americanistas (Bogotá, julio de 1985), Bogotá, Instituto Caro e Cuervo, 227–245.

Hagemeijer, Tjerk (2013), *Santome,* in: Susanne Maria Michaelis et al. (edd.), *The Survey of Pidgin & Creole Languages*, vol. 2: *Portuguese-based, Spanish-based and French-based Languages*, Oxford, Oxford University Press, 50–58.

Holm, John (1989), *Pidgins and Creoles*, vol. 2: *Reference Survey*, Cambridge, Cambridge University Press.

Holm, John (1992), *Popular Brazilian Portuguese: a semi-creole*, in: Ernesto d'Andrade/Alain Kihmn (edd.), *Actas do colóquio sobre "crioulos de base lexical portuguesa"*, Lisboa, Colibri, 37–66.

Holm, John (2004), *Languages in contact: the partial restructuring of vernaculars*, Cambridge, Cambridge University Press.

Jon-And, Anna (2011), *Variação, contato e mudança linguística em Moçambique e Cabo Verde—A concordância variável de número em sintagmas nominais do português*, Doctoral thesis, Stockholm, Stockholm University.

Kabatek, Johannes (2009), *Review of Anthony Julius Naro/Maria Marta Pereira Scherre: Origens do português brasileiro*, Revista Internacional de Lingüística Iberoamericana 13/1, 302–305.

Koefoed, Gert/vanMarle, Jap (2004), *Morphological Change I: Fundamental issues*, in: Gert E. Booij et al. (edd.), *Morphologie. Ein Internationales Handbuch zur Flexion und Wortbildung*, vol. 2, Berlin/New York, De Gruyter, 1574–1589.

Labov, William (1982), *Building on Empirical Foundations*, in: Winfred Lehmann/Yakov Malkiel (edd.), *Perspectives on Historical Linguistics*, Amsterdam, Benjamins, 17–92.

Labov, William (1994), *Principles of Linguistic Change*, Oxford/Cambridge, Blackwell.

Labov, William (2007), *Transmission and diffusion*, Language 81, 344–387.

Lefèbvre, Claire (1998), *Creole genesis and the acquisition of grammar: the case of Haitian Creole*, Cambridge, Cambridge University Press.

Lightfoot, David W. (1979), *Principles of Diachronic Syntax*, Cambridge Studies in Linguistics 23, New York/London, Cambridge University Press.

Linhares, Maria Yeda L. (^{2}2002), *A pecuária e a produção de alimentos na colônia*, in: Tamás Szmrecsányi (ed.), *História econômica do período colonial*, São Paulo, Hucitec, 109–122.

Lipski, John (2008), *Afro-Bolivian Spanish*, Madrid/Frankfurt, Iberoamericana/Vervuert.

Lucchesi, Dante (1998), *Sistema, mudança e linguagem*, Lisboa, Colibri.

Lucchesi, Dante (2001), *As duas grandes vertentes da história sociolinguística do Brasil*, DELTA 17/1, 97–130.

Lucchesi, Dante (2003), *A categoria gramatical do gênero: universais, mudança e crioulização*, in: Ivo Castro/Inês Duarte (edd.), *Razões e emoção: miscelânea de estudos em homenagem a Maria Helena Mira Mateus*, Lisboa, Imprensa Nacional-Casa da Moeda, 429–450.

Lucchesi, Dante (2004), *A participação do contato entre línguas na formação do português popular do Brasil*: novas evidências empíricas, in: Mauro Fernández/Manuel Fernández-Ferreiro/Nancy V. Veiga (edd.), *Los criollos de base ibérica*, Madrid/Frankfurt, Iberoamericana/Vervuert, 215–226.

Lucchesi, Dante (2006a), *Parâmetros sociolinguísticos do português brasileiro*, Revista da ABRALIN 5/1–2, 83–112.

Lucchesi, Dante (2006b), *Século XVIII, O século da lusofonização do Brasil*, in: Werner Thielemann (ed.), *Século das luzes: Portugal e Espanha, o Brasil e a Região do Rio da Prata*, Frankfurt, TFM, 351–370.

Lucchesi, Dante (2008), *Aspectos gramaticais do português brasileiro afetados pelo contato entre línguas: uma visão de conjunto*, in: Cláudia Roncarati/Jussara Abraçado (edd.), *Português brasileiro II: contato linguístico, heterogeneidade e história*, Niterói, EDUFF, 366–390.

Lucchesi, Dante (2009), *História do contato entre línguas no Brasil*, in: Dante Lucchesi/Alan Baxter/Ilza Ribeiro (edd.), *O português afro-brasileiro*, Salvador, EDUFBA, 41–73.
Lucchesi, Dante (2011), *Racismo linguístico ou ensino democrático e pluralista?*, Grial—Revista Galega de Cultura 190/49, 86–95.
Lucchesi, Dante (2013), *O contato entre línguas e a origem do português brasileiro*, in: Eva Gugenberger/Henrique Monteagudo/Gabriel Rei-Doval (edd.), *Contacto de linguas, hibridade, cambio: contextos, procesos e consecuencias*, Santiago de Compostela, Consello da Cultura Galega, 191–220.
Lucchesi, Dante (2015a), *Língua e sociedade partidas: a polarização sociolinguística do Brasil*, São Paulo, Contexto.
Lucchesi, Dante (2015b), *A variação na concordância verbal no português popular da cidade de Salvador*, Estudos Linguísticos e Literários 52, 166–204.
Lucchesi, Dante/Baxter, Alan (2009), *A transmissão linguística irregular*, in: Dante Lucchesi/Alan Baxter/Ilza Ribeiro (edd.), *O português afro-brasileiro*, Salvador, EDUFBA, 101–124.
Lucchesi, Dante/Baxter, Alan/Ribeiro, Ilza (edd.), *O português afro-brasileiro*, Salvador, EDUFBA, 2009.
Lucchesi, Dante/Mello, Camila (2009), *A alternância dativa no português afro-brasileiro: um processo de reestruturação original da gramática*, Papia 19, 153–184.
Lucchesi, Dante/Mendes, Elisângela (2009), *A flexão de caso dos pronomes pessoais*, in: Dante Lucchesi/Alan Baxter/Ilza Ribeiro (edd.), *O português afro-brasileiro*, Salvador, EDUFBA, 471–488.
Lumsden, John (1999), *Language acquisition and creolization*, in: Michel DeGraff (ed.), *Language creation and language change: creolization, diachrony, and development*, Cambridge, The MIT Press, 129–157.
Massini-Cagliari, Gladys (2004), *Language policy in Brazil. Monolingualism and linguistic prejudice*, Language Policy 3, 3–23.
Mattos e Silva, Rosa Virgínia (2004), *Ensaios para uma sócio-história do português brasileiro*, São Paulo, Parábola.
Mattoso, Katia ([3]2003), *Ser escravo no Brasil*, São Paulo, Brasiliense.
Megenney, William W. (1993), *Helvecian Portuguese: vernacular dialect or true creole?*, Journal of Pidgin and Creole Languages 8/2, 237–240.
Mello, Heliana (1997), *The genesis and the development of the Vernacular Brazilian Portuguese*, Doctoral thesis, New York (NY), University of New York.
Mendonça, Renato (1933), *A influência africana no português do Brasil*, Rio de Janeiro, Sauer.
Monteiro, John Manuel (1995), *Negros da terra, índios e bandeirantes nas origens de São Paulo, São Paulo*, Companhia das Letras.
Mota, Maria Antónia/Miguel, Matilde/Mendes, Amália (2012), *A concordância de P6 em português falado: os traços pronominais e os traços de concordância*, Papia 22/1, 161–188.
Mussa, Alberto (1991), *O papel das línguas africanas na história do português do Brasil*, Master's thesis, Rio de Janeiro, Universidade Federal do Rio de Janeiro.
Naro, Anthony/Scherre, Marta (1993), *Sobre as origens do português popular do Brasil,* DELTA 9, 437–454.
Naro, Anthony/Scherre, Marta (2007), *Origens do português brasileiro*, São Paulo, Parábola.
Parkvall, Mikael (2000), *Out of Africa: African influences in Atlantic creoles*, London, Battlebridge.
Parkvall, Mikael/Álvarez López, Laura (2003), *Português vernáculo brasileiro e a hipótese da semi-crioulização*, Revista da ABRALIN 2/1, 111–152.
Petter, Margarida (2006), *As línguas africanas no Brasil*, in: Suzana Cardoso/Jacyra Mota/Rosa Virgínia Mattos e Silva (edd.), *Quinhentos anos de história linguística do Brasil*, Salvador, Secretaria da Cultura e Turismo do Estado da Bahia, 117–141.

Ribeiro, Darcy (1995), *O povo brasileiro: evolução e sentido do Brasil*, São Paulo, Companhia das Letras.

Ribeiro, João ([3]1906, [1]1889), *Diccionario grammatical*, Rio de Janeiro, Francisco Alves, 1906.

Rickford, John (1987), *Dimensions of a creole continuum: history, texts and linguistic analysis of Guyanese Creole*, Stanford, Stanford University Press.

Rodrigues, Aryon (1986), *Línguas brasileiras: para o conhecimento das línguas indígenas*, São Paulo, Loyola.

Rodrigues, Aryon (2006), *As outras línguas da colonização do Brasil*, in: Suzana Cardoso/Jacyra Mota/Rosa Virgínia Mattos e Silva (edd.), *Quinhentos anos de história linguística do Brasil*, Salvador, Secretaria da Cultura e Turismo do Estado da Bahia, 143–161.

Rougé, Jean-Louis (2008), *A inexistência de crioulo no Brasil*, in: José Luiz Fiorin/Margarida Petter (edd.), *África no Brasil: a formação da língua portuguesa*, São Paulo, Contexto, 63–74.

Sapir, Edward (1921), *Language: an introduction to the study of speech*, New York, Harcourt, Brace and company.

Saussure, Ferdinand de (1916), *Cours de linguistique générale*, Lausanne/Paris, Payot.

Scherre, Maria Marta (1988), *Reanálise da concordância nominal em português*, Doctoral thesis, Rio de Janeiro, Universidade Federal do Rio de Janeiro.

Siegel, Jeff (2006), *Links between SLA and Creole Studies. Past and Present*, in: Claire Lefèbvre/Lydia White/Christine Jourdan (edd.), *L2 Acquisition and Creole Genesis—Dialogues*, Amsterdam/Philadelphia, Benjamins, 15–46.

Siegel, Jeff (2008), *The Emergence of Pidgin and Creole Languages*, Oxford, Oxford University Press.

Silva Neto, Serafim da ([2]1963, [1]1951), *Introdução ao estudo da língua portuguesa no Brasil*, Rio de Janeiro, INL.

Silva Neto, Serafim da ([5]1988, [1]1957), *História da língua portuguesa*, Rio de Janeiro, Presença.

Trudgill, Peter (2009), *Sociolinguistic typology and complexification*, in: Geoffrey Sampson/David Gil/Peter Trudgill (edd.), *Language complexity as an evolving variable*, Oxford, Oxford University Press, 97–108

Trudgill, Peter (2010), *Contact and Sociolinguistic Typology*, in: Raymond Hickey (ed.), *The Handbook of Language Contact*, Chichester, Wiley-Blackwell, 299–319.

Vieira, Sílvia Rodrigues/Bazenga, Aline (2013), *Patterns of third person plural verbal agreement*, Journal of Portuguese Linguistics 12/2, 7–50.

Winford, Donald (2003), *An Introduction to Contact Linguistics*, Amsterdam/Philadelphia, Benjamins.

Volker Noll

4 Historical phonetics and phonology

Abstract: This chapter traces the emergence of the major phonetic and phonological differences between European (EP) and Brazilian Portuguese (BP), following the evolution of these features on both sides of the Atlantic and presenting the results in sound charts.

Keywords: Brazilian Portuguese, European Portuguese, pronunciation, phonetics, phonology, diachrony, archaisms, innovation, *chiamento*, sound charts

1 General settings

The history of Spanish and Portuguese shows that the spread of a language to new territories is a strong impetus to linguistic variation and change. While the first domain subject to modification is the vocabulary, changes in pronunciation are the most noticeable. In the late Middle Ages, European Portuguese (EP) had already developed a certain written standard. In the 16th century, the grammars of Fernão de Oliveira (1536) and João de Barros (1540) were published with some information on phonetics. The emergence of Brazilian Portuguese (BP) pronunciation which was subsequent to discovery (1500) and the beginning of settlement (1532), however, did not spring from a specific EP accent like the one in Lisbon. On the contrary, settlers from all over the mother country from between the Minho River and the Algarve met in different contexts which led to dialect levelling and the formation of regional koines in Brazil.

Between 1532 and 1600, 17 towns were founded along the vast Brazilian coastline between Cananéia (São Paulo) and Natal (Rio Grande do Norte). With the foundation of São Luís (1612) and Belém (1616), the extension of territories to be colonized reached the Amazon Delta in the North. In spite of Salvador being a metropolis and the first capital from 1549 to 1763, the crucial external factors driving the evolution of BP were the size of the country, the persistent isolation of areas due to the lack of overland routes, contact with other languages and a cultural life which was without printing until the 19th century.

The purpose of this chapter is to retrace the development of BP pronunciation, specifically in comparison with the linguistic situation in Portugal from the 16th century to present. The procedure is based on early linguistic sources showing (archaic) Brazilian features which are no longer part of standard EP today (see Sections 2, 3.1) as well as others which newly formed in Brazil (see Section 3.2). After shedding some light on the question of possible regional or specific EP influence on the constitution of BP pronunciation (see Section 4), we will classify our findings and present the structural differences between EP and BP due to diachronic evolution in sound charts

https://doi.org/10.1515/9783110405958-005

(see Section 5). In spite of numerous case studies on various aspects of EP or BP phonetics individually having been published, articles on the evolution of BP phonetics in comparison with EP are scarce (cf. Révah 1958; Noll 2008, 219–243).

2 BP pronunciation in early linguistic sources

Although early written sources from Brazil are better documented nowadays than they once were, it is still quite difficult to establish a reliable chronology of phonetic changes in BP, as there is hardly any material available from before the 18th century. If we include the history of EP, that is to say, cases where linguistic contrast depends on change in EP, the task becomes more manageable.

The first difference between EP and BP phonetics was casually documented in the 16th century. In 1569, Welsh physician John David Rhys described the pronunciation of the newly formed fricative allophone /d/ [ð] in Spanish and EP as weak and compared it to the sound in Greek ("Laxè ut plurimùm hanc literam efferunt Hispani, ac Lusitani [...] In his enim atque similibus D Grecorum δ sono emulatur"; 1992, 118). Only a few years later, Duarte Nunes de Leão confirmed this observation, noting that /d/ was pronounced "with the tongue between the teeth" (1576, 57). As Brazilian voiced plosives (/b/, /d/, /g/) are never fricative, as in EP *roda* [ð] (BP [d]), the emerging difference between the two varieties of Portuguese became noticeable within the first century after discovery.

The first hint of change in Brazilian pronunciation comes from Frei Vicente do Salvador in his *História do Brasil* in 1627. In a scene describing an attack by the Dutch in northeastern Pernambuco, he quotes a black slave with the following words: "Não retira, não, sipanta, sipanta" (1627, 370; 'Don't retreat, beat off the enemy!'). This is a testimony of metathesis in substandard vernacular BP as a result of adjusting the syllable structure (VC > CV) of the regular form *espanta*.

The first purposeful metalinguistic remark on BP pronunciation dates from the middle of the 18th century. In *Compendio de orthographia*, Monte Carmelo criticizes Brazilian speakers for confounding pretonic vowels like in "Prégár" [ɛ] 'to preach' and "Pregar" [ə] 'to nail' (1767, 128, 131). Brazilians pronounce both words alike, either with an open-mid [ɛ] or with a close-mid [e], partly depending on their regional background.

A female slave's handwritten letter from Piauí in 1770 reveals additional novel features of BP pronunciation (cf. Mott 1979). The document, signed Esperança Garcia, gives evidence for vowel epenthesis highly typical of spoken BP in learned consonant clusters (*adm**i**nistrar* → <adeministrar> [i]). A less common way to achieve this change, documented in the letter as well, is metathesis (*pr**o**curador* → <Porcurador>). Moreover, the spelling <abacho> (*abaixo*) represents the monophthongization of <ai> and <ei> before a subsequent prepalatal sibilant, [ʃ] or [ʒ], in colloquial BP. A few years later, in 1798, the statutes of two episcopal seminaries that oversaw the education of or-

phans in Olinda and Recife, once again in Brazil's Northeast, refer to common language mistakes to be avoided in class (cf. Azeredo Coutinho 1798a, 1798b). The erroneous spelling of final <-e> as <-i> (<febri>) represents today's standard Brazilian pronunciation of <-e> [-i] in this position. EP pronounced <-e> [-i] alike until the beginning of the 18th century, but then reduced this final vowel to [-ə] throughout that century. Another innovation documented in the statutes and now common in spoken BP is the monophthongization of <ei> (<Janero> < *janeiro*; still, there are some exceptions: BP *jeito* [e̯i̯], *lei* [e̯i̯]; see Section 3.2). Standard EP did not maintain this diphthong either, and changed it into [ɐ̯i̯] in the 19th century (EP [ʒɐˈnɐ̯i̯ru]).

Around 1800, a play, *O periquito ao ar, ou O velho uzurário* by Manuel Rodrigues Maia (ca. 1800), in which a rich Brazilian going to get married in Portugal is imitated through his particular linguistic habits, confirms final <-e> [-i] as a Brazilian pronunciation at that time (<mi> *me*). Moreover, the personal pronoun *lhe* spelled <le>, <li>, testifies to the depalatalized form [li] BP currently employs (see Noll 2008, 159–165).

This short historic overview shows that, at the beginning of the 19th century, some major phonetic features of today's BP had not yet showed up in texts. Among these are the affrication of /t/ and /d/ in front of [i] (*tio* ['tʃiu̯], *dia* ['dʒia]), the vocalization of implosive /l/ (*bolso* [ou̯], *Brasil* [iu̯]), and the intrusion of an epenthetic [i̯] in final stressed syllables ending in /s/ (*atrás* [a'trai̯s]).

3 Archaism and innovation in BP

Two major points account for the current differences in phonology and phonetics between EP and BP noticed today. On the one hand, changes occurred in EP. In this case, Brazilian features are considered archaisms, given that the point of reference usually is the area when the language started its expansion. Archaic features mainly concern Brazilian vocalism which preserved major traits of Portuguese vocalism up to the 18th century, whereas EP weakened unstressed vowels during that century. On the other hand, changes are due to linguistic evolution on the Brazilian side. These innovations mainly concern consonants and the formation of epenthetic vowels in BP.

In addition to this general setting, there have been numerous attempts to relate nearly all the salient features of BP pronunciation almost indiscriminately to both indigenous and African influence. It would be beyond the scope of this chapter to discuss them in detail. However, if we try to contrast indigenous and African stratic implications in BP with a typical regional focus, no crucial differences can be detected between the historically Tupi-influenced hinterland of São Paulo in the Southeast and the language which emerged in interaction with African slaves in the sugar cane regions of Bahia and Pernambuco in the Northeast (*Região Nordeste*). The impact of linguistic convergence or multiple causation cannot be ruled out though. The same is true for linguistic interference ascribed to some influence of varieties from southern

Portugal in particular, as well as possible linguistic consequences of regional immigration from the Azores in southern Brazil (cf. Noll 2008, ch. 6.5, 9.2, 9.3).

3.1 Archaic features in BP

The linguistic notion of archaism signifies that a specific feature in a speech area has not followed the same development as that of another region, usually a linguistic center belonging to the same variety, language or family of languages; it does not imply any evaluative assessment. In polycentric languages like Portuguese, the point of reference is the genuine speech area or standard language, depending on the state of diachronic linguistic evolution. Therefore, present standard EP pronunciation serves as a reference framework for the classification of phonetic characteristics of BP. The following features of Brazilian pronunciation reflect earlier states of language in Portugal. Chronological overlaps and divergent regional developments aside, the characteristics of contemporary BP mentioned below are typical of Portuguese until the centuries indicated.

16th and 17th centuries

- *Voiced plosives*: Voiced plosives ([b], [d], [g]) have been maintained in BP (*cansa-d̲o* [kɐ̃'sadu]), while they produced fricative allophones ([β], [ð], [ɣ]) in central and northern Portugal. This is especially conspicuous for /d/. The change in EP occurred in the 16th century, except word-initially and following a nasalized vowel (EP *cansad̲o* [kɐ̃'saðu]; *quando* ['kwɐ̃du]).
- *Heterosyllabic nasalization*: BP has preserved heterosyllabic nasalization mainly in stressed syllables (cf. Moraes/Wetzels 1992, 153–154), which means that the triggering nasal consonant is part of the following syllable (*c̲ama* ['kɐ̃.ma]). The intensity of this kind of nasalization varies and is most typical of Brazil's Northeast. In EP, heterosyllabic nasalization was lost in the 16th and 17th centuries (['kɐ.ma]), aside from regional occurrence in the North and the South. Nasalization in standard EP is only tautosyllabic (*canto* ['kɐ̃.tu]), as in present day French.
- *Diphthong* <ai>: BP has partially maintained the 16th century variation in the diphthong <ai> which may be reduced to [a] in front of prepalatal [ʃ] (*baixo* ['baʃo]). In Camões' epos *Os Lusíadas*, we find alternating forms such as *baixo—baxo* and *debaixo—debaxo*, with the diphthongized forms being more frequent. In BP, this variation (*abaxo*) is documented in the previously mentioned letter from 1770 (Mott 1979, 8; see Sections 2, 3.1). In present day EP, monophthongization is limited to the southern part of the country.
- *No opposition of* /ɐ/ *vs.* /a/: BP never developed the phonological opposition of /ɐ/ vs. /a/, which marks the difference between the present tense and the simple

perfect (*pretérito passado simples*) in the first person plural of the first conjugation in EP standard (*cant<u>a</u>mos* vs. *cant<u>á</u>mos*). In the 16th century, [a] and [ɐ] were mostly allophones, but João de Barros marks the perfect of *amar* with an accent as early as 1540 ("Amámos"), which represents the so-called *a grande* [a] (Barros 1540, 22). Nonetheless, Teyssier (1966, 143) and Azevedo Maia (1986, 315) argue that the etymologically unfounded opposition between the two vowels had not yet been fully established in the 16th century. In Portugal's northern dialects, both forms coincide in [a], whereas they appear as [ɐ] in the South nowadays.

- *Diphthong* [o̯u̯]: BP has preserved the diphthong [o̯u̯] (< Lat. <au>) as a free variant for <ou> (*vou* [vo̯u̯], [vo]; *amou* [a'mo̯u̯], [a'mo]). However, no variation between <ou> and <oi>, as in EP pairs like *touro*, *toiro*, exists. In standard EP and the south of Portugal, [o̯u̯] was monophthongized in 17th century. In BP, the use of the diphthong is also linked to emphasis.
- *Medieval affricate* [tʃ]: In the southern part of Brazil (the states of São Paulo, Minas Gerais, Paraná, and especially Mato Grosso), the medieval affricate [tʃ] <ch> (*chave* [tʃ]) has been maintained regionally as in northern Portugal. In central and southern Portugal, the sound lost its plosive component in the 17th century.
- *Final vowels*: In rural areas of São Paulo and southern Brazil (Paraná, Santa Catarina, and Rio Grande do Sul), final [-e], [-o] (*sete* [-e]) represent Portuguese vocalism, before EP final vowels were weakened in the 18th century (cf. Noll 2008, 56, n. 15).
- *Pretonic variation*: BP has maintained a certain variation in pretonic close-mid and close vowels ([e - i], [ẽ – ĩ], [o – u]), all of which have been stable in EP since the 18th century. Metaphonic vowel raising is triggered by a subsequent high vowel (cf. BP *m<u>e</u>nino* [mi'ninu]). A historic example is *d<u>o</u>rmir* (BP [dux'mi^{x}]) quoted in Portugal by Fernão de Oliveira in 1536: "[...] quase nos confundimos, dizendo uns *somir* e outos *sumir* e *dormir* ou *durmir* [...]" (1975, 64). In *Historia Naturalis Brasiliae*, Markgraf/Piso note the sloth (*preguiça*) as "*Priguiza*" (1648, 221), which reflects the pretonic [i] of the spoken language. Metaphony, as seen here, is a common phenomenon in Romance languages (cf. Lat. *fēcī* → Port. *fiz*).
- *Basic quality of pretonic vowels*: Except for cases of metaphonic interference, BP tends to preserve the basic quality of pretonic vowels in [e ɛ], [a], and [o ɔ] (*m<u>e</u>temos* [me'temus], *f<u>a</u>lamos* [fa'lɐmos], *m<u>o</u>rango* [mo'rẽgu]), which, in many cases, were centralized/raised to [ə], [ɐ], and [u] in EP in the 18th century (for further development, see Section 3.2).
- *Alveolar* /s/ *and* /z/: Much of Brazil has preserved alveolar /s/ and /z/ in implosive position (*festa* ['fɛsta], *nascimento* [nasi'mẽtu], *mesmo* ['mezmu]). These sibilants were palatalized in EP mainly during the 18th century (*festa* ['fɛʃtɐ], *nascimento* [nɐʃsi'mẽtu], and *mesmo* ['meʒmu]). As the linguistic situation in Brazil is fairly complex, it will be examined in Section 4).

18th and 19th centuries

- *Final* [-a], [-i]: In BP, the vowels [-a] and [-i] in unstressed final position (*casa* ['kaza], *lume* ['lumi]) reflect the state of Portuguese at the beginning of the 18th century, as standard EP changed these final vowels to [-ɐ] and [-ə] throughout that century. Remnants of final [-i] can still be found in northern and southern Portugal. In current BP, there is a noticeable tendency to weaken final /a/ to [-ɐ], thus following the evolution of EP.
- [e] + *palatal*: BP preserves the pronunciation of [e] followed by the subsequent palatals [ʎ ɲ ʃ ʒ] (*espelho*, *venho*, *vejo*, *fecho* [e]). Starting from Lisbon, EP centralized this vowel in the 19th century with different results ([iʃ'pɛʎu], ['vɐɲu], ['vɐ(i̯)ʒu], ['fɐi̯ʃu]). The Brazilian linguist José Paranhos da Silva humorously commented on this: "*Paixão* significa em Portugal *peixe grande*" 'In Portugal, *paixão* ['passion'] means *big fish*' (1879, II, 8; cf. EP *peixão* [ɐi̯]).
- *Diphthongs* [ei̯], [ẽi̯]: Similar to the conservation of [e] in contact with palatal sounds, BP maintains the diphthongs [ei̯] and [ẽi̯] (*jeito* ['ʒei̯tu], *bem* [bẽi̯]) which, starting from Lisbon, was centralized in EP in the 19th century to become [ɐi̯] and [ɐ̃i̯], respectively (['ʒɐi̯tu], [bɐ̃i̯]).

3.2 Innovative features in BP

Until the 1990s, little was known about the internal history of Portuguese in Brazil. Printing was not available until the 19th century. When slavery was abolished in 1888, many local documents were destroyed. Linguistic research in Brazilian archives has mainly been undertaken in the last two decades. Nonetheless, records allowing the close following of the evolution of phonetic features of BP are quite scarce. They do not reach beyond the 18th century and are not always explicit.

18th century

- *Pretonic vowels*: As explained in the Section 3.1, BP tends to maintain the basic quality of pretonic vowels, while EP weakened them in the 18th century. As pretonic variation is multifaceted, BP has also undergone changes in this respect. In the 16th century, Portuguese exhibited eight units in pretonic position [i e ɛ ɐ a ɔ o u]. BP eliminated [ɐ] and opted either for close-mid [e], [o] or open-mid vowels [ɛ], [ɔ], which led to two alternative sets of five units: [i e a o u] vs. [i ɛ a ɔ u]. Originally, BP used to be classified in two zones according to the degree of aperture in pretonic <e> and <o>, the open pronunciation being ascribed to the regions north of a line reaching from the Bolivian border in the west, through Cuiabá and

Goiânia, to the estuary of the Rio Mucuri in southern Bahia (Nascentes 1953, 25–26). However, the new *Altas Linguístico do Brasil* (ALiB) shows that the distribution described above is merely approximate. There is only a tendency towards vowel aperture in the Northeast (*Região Nordeste*: states from Bahia to Maranhão) without excluding either close-mid or open-mid vowels in the rest of the country (cf. ALiB, cartas F01 V 1–V 4G).

Contemporary EP has eight pretonic units [i ɛ a ə ɐ ɔ o u], as it lost [e'] when reducing pretonic [e], [a], [o] to [ə], [ɐ], [u] in the 18th century. Still, the original vowel aperture has been maintained in consonant clusters (EP *acção* [a'sɛ̃ũ̯]), before implosive /l/ (EP *relvado* [rɛɬ'vaðu]), and in vowels originating from former hiatuses (Lat. *praedicare* > *pre-egar* > *pregar* [ɛ'] 'to preach'; Lat. *colorare* > *corar* [ɔ']).

Vowel raising due to metaphony is another factor that interfered with pretonic vowel quality in BP ([e'], [o'] → [i'], [u']). This does not necessarily reflect pretonic variation of former centuries. Vowel raising in BP may be caused when /o/ comes in contact with labial consonants like [p b f v m] (*boneca* [bu'-]).

The vacillation of Brazilian pretonic vowels is a result of the evolution of mainly spoken language over a long time, and it carries certain subtleties. While *sotaque* is always pronounced [su'-], *comer* [ku'-] would be clearly informal, whereas *morango* with pretonic [u] would not be considered a Brazilian pronunciation.

- *Consonant clusters*: Consonant clusters in words of learned origin (plosive + fricative/nasal) are broken up in BP by inserting an epenthetic [i] (*adv̲ogado* [adʒivo'gadu]). The 1770 letter from a female slave produces evidence for this in "*ademinstrar*" (cf. Mott 1979, 8). The objective is to adjust the syllable structure VC. CV to the basic pattern of CV.CV. Words ending in a consonant other than the ones usually admitted by the phonotactics of Portuguese also take a paragogic vowel (CVC → CV.CV). This concerns mostly loanwords (*tíquete* [-tʃi]), proper names (*Maluf* [-fi]), and acronyms (VARIG [-gi]). As EP reduced final /e/ to [-ə] in the 18th century, which quite often becomes silent (cf. EP *dirigente* [dəri'ʒẽt(ə)]), the tolerance for final CVC is higher in EP than in BP.
- *Monophthongization of* <ei>: The monophthongization of <ei> [ei̯] > [e] is very common in BP and has partly turned the two realizations into free variants. The diphthong is usually preserved in stressed final position (*amei*), in front of another vowel (*passeio*) or /t/ (*jeito*), and following prepalatals /ʒ ʃ/. The suffixes *-eiro*, *-eira* favor monophthongization. In Brazil, the change was mentioned for the first time at the end of the 18th century in "*Janero* em lugar de *Janeiro*" (Azeredo Coutinho 1798a, 47; Noll 2008, 170). In EP, it appears in the Alentejo region and was documented only a few years earlier in 1769 (*sardenhero* for *sardinheiro*; cf. Teyssier 1984, 64, 107).
- *Dephonologization of* /ʎ/: In BP, we observe the partial dephonologization of /ʎ/ (→ [l], [j]). In the object pronoun *lhe* [li], it is general and was documented around 1800 in the play *O periquito ao ar, ou O velho uzurário* by Manuel Rodrigues Maia (<li>, cf. Noll 2008, 162). The extended use of depalatalized /ʎ/ [l] is common in

colloquial substandard BP (*mulher* [mu'lɛ]). The delateralized variant [j] is more marked socially. Early examples ("[...] *teado* em lugar de *telhado*: *fio* em lugar de *filho*") are given in the instructions for teaching orphans in Recife and Olinda in 1798 (cf. Azeredo Coutinho 1798a; 1798b; Noll 2008, 170–171).

– /e/ + *nasal*: The EP opposition of mid-open and mid-close [ɛ e], [ɔ o] is neutralized in BP in front of a nasal consonant (*prémio*, EP [ɛ], BP [e]). In BP, this represents an extension of vowel closure, since Portuguese already closed /a/ [ɐ] in this position. Fernão de Oliveira gives evidence for [ɐ] in EP preceding a nasal consonant in 1536 ("[...] *a* pequeno, como *Alemanha*", Oliveira 1975, 48). There is currently no evidence for when exactly the closure of [ɛ] and [ɔ] occurred in BP.

19th and 20th centuries

– *Vocalization of implosive* /l/: Given the velarized nature of /l/ in Portuguese, BP turned it into a semivowel in implosive position (*bolso* [oʊ̯], *Brasil* [iʊ̯]), which is historically common within the Romance languages. The vocalization of /l/ also appears in EP in the dialect of Alto Minho. In BP, the vocalized pronunciation is part of the non-defined standard, and the transcription in BP dictionaries occasionally marks it. Implosive velarized /l/ [ɬ] only appears locally in, for example, the speech of elderly individuals in Rio Grande do Sul. There may be hint of early vocalization in the *Viola de Lereno*, a collection of Brazilian *modinhas* by Domingos Caldas Barbosa (1826, II, 19):

  Se não tens mais quem te sirva
  O teu moleque sou eu
  Chegadinho do Brasil
  Aqui stá que todo é teu.

  In analyzing the rhyme of *eu*, *teu* and *Brasil* [-iu] (?) it may be tempting to suppose vocalization of /l/, but it remains uncertain. Antenor Nascentes (1953, 48) gives reference to the subtle pronunciation of the final /l/ by Rio's upper class in the first quarter of the 20th century. Therefore, it can be assumed that the process started in the 19th century.
– *Affrication of* /t/, /d/: The affrication of /t/ and /d/ in front of [i] (*tio* ['tʃiʊ̯], *dia* ['dʒia]) is probably the most salient feature of BP pronunciation and is part of the non-defined standard. Nonetheless, there are places in the interior (rural zones, especially in the South) and large areas in Brazil's Northeast which preserve the original plosive articulation. These areas stretch from rural Bahia (excluding the capital Salvador), Sergipe, Alagoas, Rio Grande do Norte (including the three capitals) to rural Ceará. Ceará's capital Fortaleza, however, exhibits affrication. This distribution clearly marks affrication as an urban evolution. None of the 19th cen-

tury writings on grammar mentions it at that time. In the 1940s, Brazilian teachers tried to eradicate affrication (Stavrou 1947, 26). Transcription in BP dictionaries does not usually mark it. Considering its vast expansion nowadays, we may assume the process started in the 19th century. At present, the evolution has taken a further step, as some speakers tend to fade out [i] in final position (*noite* → [noitʃ]).

– *Epenthesis of* [i̯]: Another prominent feature of BP is the epenthesis of the semivowel [i̯] in front of /s/ in a stressed final syllable (*atrás* [a'traiˆs]). São Paulo, Mato Grosso do Sul, and the three southern states of Paraná, Santa Catarina, and Rio Grande do Sul tend to avoid epenthesis, although it occasionally appears there as well. Azeredo Coutinho does not mention this pronunciation in 1798, but it is apparent in the spelling of *mês* in a document written by a person of African descent in Bahia in 1841: "Ao sete dias do Meis de Novembro [...]" (Oliveira 2003, II, 405). A few years later, in 1848, poet Braz Pitorras from Ceará rhymes *mais* and *me dás* [daiˆs], providing further evidence for the phenomenon's existence at that time (Seraine 1949, 62):

> Minha Ignês, não posso mais
> Tanto silêncio guardar
> Novas tuas não me dás.

– *Velarization and glottalization of rhotics*: In BP, formerly trilled /r̄/ in <rr>, <r-> was velarized and reduced to the glottal spirant [h] (*carro*, BP ['kahu]; *rio*, BP [hiu̯]). A similar process started in Lisbon in the 19th century. The uvular fricative [ʁ] now mainly covers central and southern Portugal (['kaʁu], [ʁiu̯]), while the North has maintained [r̄]. In implosive position, /r/ is a simple flap in EP. Nowadays, BP mainly preserves flapped /r/ only in intervocalic position. In preconsonantal position, /r/ becomes a velar or glottal fricative (*quarto* ['kwaxtu], ['kwahtu]). The situation becomes quite complex in São Paulo, the South and some interior zones, however. In these areas, [r̄] and [r] have been maintained in changing conditions, although velarization proceeds (Paraná, Santa Catarina). Furthermore, preconsonantal /r/ can also become retroflex (['kwaɽtu]), known as the so-called *r-caipira*, referring to its rural origin. Retroflexion is common in southern Brazil, the state and the city of São Paulo, Mato Grosso do Sul, Mato Grosso, Goiás, and the south of Minas Gerais. In final position, /-r/ in BP becomes silent in many cases. In Rio de Janeiro, it is often a velar fricative, in São Paulo and the South often a flap or retroflex (especially in Mato Grosso do Sul and Mato Grosso).
The beginning of velarization is documented as follows. Shortly after Brazil's independence, the Visconde de Pedra Branca, a Brazilian diplomat, was asked to give some comparative information on BP to be published in the *Introduction à l'Atlas ethnographique du globe* in Paris. The explanation given for EP "cecia" surprisingly refers to the uvular or velar pronunciation of "r" ('Action de grasseyer' 'to pronounce the uvular ‹‹r››', Balbi 1826, 173). This may be the first description of

the phenomenon in Portuguese. Gonçalves Viana (1883, 48) stated that Brazilians frequently pronounced initial "r" as a voiced fricative. The glottal realization of /r̄/, /r-/ [h] in BP was documented by Stavrou (1947, 30) who compared the pronunciation in Rio de Janeiro to *how* and *Harry* in English.
The evolution of rhotics has gone further in BP than in EP. In Brazil, the former vibrant has turned into a velar and even a glottal consonant. The change occurred in initial position precedes the preconsonantal change, the latter likely having started in the 19th or 20th century. In final position, /r/ is very faint in colloquial BP and tends to become silent, especially in infinitives. Padre Lopes Gama from Recife criticizes this speech habit as early as 1842: "Muitos declarão guerra aos *rr* finaes, e dizem sempre *mandà*, *buscà*, *comê*, *dormí*, *singulà*, &c. &c." (facsimile, cf. Pessoa 1994, 78). The beginning of retroflexion (*r-caipira*) is undocumented. However, it is certain that this process has nothing to do with the Tupi substrate, as Tupi has a simple alveolar flapped /ɾ/.

– Apart from systematic changes, there are special cases like the stressed vowel in BP *senhora* [siˈɲɔra], with an open [ɔ] as opposed to EP [səˈɲoɾɐ]. Lopes Gama from Recife mentions it in his satirical newspaper *O Carapuceiro* in 1842: "[...] mas de senhor formão *senhóra* com *o* bem aberto" (facsimile, cf. Pessoa 1994, 78).

4 Regional EP influence on emerging BP pronunciation?

The majority of archaic features in BP mentioned above differ mainly from standard EP after the fundamental changes in the 18th and 19th centuries had taken place. With regard to the alleged role of Andalusian pronunciation in the history of Latin American Spanish, linguists wondered if some major regional EP influence possibly contributed to the formation of BP phonetics. Among others, William Entwistle (1982, 377) believed that BP originated from the variety of the Algarve. This idea, however, may be a misconception in light of Portuguese dialectology. Unlike southern Spanish from the part of Andalusia that faces the Atlantic Ocean, southern characteristics of Portuguese extend to over half of Portugal. In former times, Portuguese varieties from the very south up to the Mondego River were classified as southern. In fact, the southern and central varieties (*falares centro-meridionais*), as we call them today, mainly contrast with the area of northern Portuguese north and east of Aveiro (Minho, Trás-os-Montes, Alto Douro, Beira Alta). Southern features in the stricter sense as those of the Algarve are certainly not typical of BP (cf. Kröll 1994, 551).

The formation of a koine or of koines which was subsequent to the expansion of different regional varieties from Spain and Portugal in Latin American territories is always based on the levelling of prominent dialect features. This process depends on the speakers' acceptance or rejection, but not necessarily on the number of immi-

grants from a certain region. It is obvious that specific features of northern Portuguese, such as the apico-alveolar sibilants /s̺/, /z̺/ or the merger of /b/ and /v/, never appeared in BP (apart from occasional variants like *bravo*/*brabo*, *vassoura*/*bassoura*). Therefore, BP rather followed the southern and central varieties of EP. Other peculiarities like today's northern [oʊ̯] (*ouro*) and [tʃ] (*chave*) as well as final [-i] (*leite*) in southern and in northern Portugal are historic forms and represent formerly widespread features in Portugal.

The innovative features of BP discussed above cannot be connected to any specific regional influence from Portugal. The only existing similarity is the monophthongization of <ei> [eɪ̯] > [e] in Brazil and in the southern half of Portugal, which dates back to the 18th century. This change started rather simultaneously on both sides of the Atlantic, though without being complete in Brazil (cf. *-eiro*, BP [ero], [eɪ̯ru]). Another phonetic change which ran parallel is the velarization/glottalization of rhotics. It may have started earlier in BP, as the evolution there has reached a more advanced level. The velarization of rhotics is quite common in different European languages (e.g., German, French, Dutch).

4.1 The special case of *chiamento*

Aside from any suppositions concerning regional EP influence on BP, there was a major event in the history of both Portugal and Brazil that gave rise to linguistic speculation. When Napoleon invaded Portugal in the Peninsular War (1807–1814), the Portuguese Court, about 15.000 people in total, moved to Rio de Janeiro and stayed there from 1808 until 1821.

The linguistic background is EP palatalization of preconsonantal and final /s/ > [ʃ] (or [ʒ] in front of voiced consonants) called *chiamento* (*mais* [maɪ̯ʃ], *mesmo* ['meʒmu]). This process likely started at the end of the 17th century in southern Portugal (cf. Silva Neto 1986, 160). Luís António Verney, who left the country in 1736, describes *chiamento* as a common feature of EP in his *Verdadeiro método de estudar* (1748, I, 29). This pronunciation is also known for being typical of Rio de Janeiro, especially during the 20th century when the knowledge of BP pronunciation was mainly limited to the situation in Brazil's former capital.

Consequently, *chiamento* in Rio de Janeiro was attributed to the presence of the Portuguese Court during those 13 years of emigration at the beginning of the 19th century. A certain number of linguists claimed its appearance to have at least some connection to EP pronunciation, if not being an outright case of dialect imitation (cf. Révah 1958, 390; Lipski 1975, 1976; Teyssier 1984, 77; Giangola 2001, Leite/Callou 2002, 32; M. Azevedo 2005, 220). This explanation represents one of the classic and persisting views regarding the history of BP. Surprisingly, none of the linguists dealing with the question found it necessary to investigate the testimonies of *chiamento* in Brazil previous to the 20th century or to determine the actual spread of *chiamento* in Brazil at present.

Would the short-lived prestige of the Portuguese Court in Rio de Janeiro until 1821 have caused the pronunciation change of implosive /s/ in the local population? Lipski's reflection on "dialect imitation" (1975, 222) suffers an early setback if we consider the judgement of the Brazilian ambassador to Paris, the Visconde de Pedra Branca, delivered in 1826. With regard to EP, Pedra Branca speaks of coarse pronunciation ("l'âpreté dans la prononciation") and the arrogance of expressions ("l'arrogance des expressions") (Balbi 1826, 173). In Brazil, the Englishman Alexander Caldcleugh, who lived in Rio de Janeiro between 1819 and 1821, comments on Brazilian pronunciation as follows: "The pronunciation of the Brazilians is not so nasal nor so jewish in the sound of the s, and on the whole it is a more agreeable language than in the mouth of a native [of the mother country]" (Caldcleugh 1825, I, 66). "Jewish" here refers to the [ʃ] typical of EP and sometimes also called *chiante mourisca* 'Moorish sibilant'. Caldcleugh does not mention *chiamento* in Brazil's capital during that time.

In 1879, the Brazilian linguist José Paranhos da Silva clearly denies in his book *O idioma do hodierno Portugal comparado com o do Brazil* that /s/ in Brazil was ever pronounced in the palatalized manner of Portugal ("[...] a pronuncia de *s*, cujo valôr no Brazil nunca foi o de *x* [...] Os Portuguezes dizem — *faxto*, *mixterio*, *livrox novox* em vez de *fasto*, *misterio*, *livros novos*", Paranhos da Silva, 1879, 20). In his *Sistema de ortographia brazileira*, which was also published in Rio de Janeiro, he even rates this pronunciation as unbearable to Brazilian ears ('a real scourge'), especially if every /s/ were to become [ʃ] (cf. § 11 below):

> [...] mas da pronùncia mourisca, leem *ech-césso*, *ech-céto*, *ech-citar*; e què nós os Brazilèiros lemos «*eşeso*, *eşeto*, *eşitar*».
> § 11. Óra, şeria uma verdadèira calamidade para os ouvidos brazilèiros què todos eşes valores şe reduzişem a o şom de chiante mourisca; o coál şó póde ter logar de vez en coàndo, como a dişonància na mùzica. Purtanto, escreveremos com *x* ùnica-mente o şom de chiante fórte, o som què şe ouve en *roxo*, *coxo*, *xadrez* (Paranhos da Silva 1880, 23–24).

Had *chiamento* been a prominent phonetic feature of Brazil's capital, this would be quite a surprising statement. In 1921, Nascentes finally gives an illuminating description of the pronunciation of /s/. According to his observations, only the upper class exhibited *chiamento* in final /s/ in Rio de Janeiro:

> As classes cultas pronunciam o *s* final, mudando entretanto numa chiante, como no Sul de Portugal. Ha quem attribua esta pronuncia ao influxo portuguez, sem explicação maior (Nascentes 1921, 317).
> 'The members of the upper class now change the final *s* into a hissing sound like in southern Portugal. Some people ascribe this pronunciation to Portuguese influence without going into further detail'.

This statement implies that *chiamento* in Rio was only partial in 1921, although the Portuguese model from a hundred years before appeared in all positions. Additionally, the change was not even common among the population of Rio de Janeiro. Under

these circumstances, it is rather unlikely that the Portuguese provided the original impetus for *chiamento* in Rio de Janeiro in 1808.

The situation becomes even clearer when analyzing the extension of *chiamento* throughout Brazil. Southern Brazil, as well as Minas Gerais and Goiás, mainly displays the historic /s/ [s z] pronunciation. Further north, there is a large intermediate zone between Bahia and Maranhão where [s z] and [ʃ ʒ] tend to be free variants in preconsonantal position within a word (mainly in contact with plosives as well as /l/ and /n/). Sometimes, the variation also occurs in final position, especially in the town of Recife (Pernambuco). We call this kind of distribution partial *chiamento*. Finally, there are isolated areas of more or less complete *chiamento* like Santos (São Paulo) as well as a part of the coastline in Santa Catarina. It comes as a surprise that complete *chiamento* (like in Rio de Janeiro) occurs in Belém do Pará, a town of about 1.5 million inhabitants in the estuarine area of the Amazon in the North. This was completely unknown in linguistic literature on BP until the publication of an article by Noll (1996). A reduced, but still predominant form of *chiamento* exists in the region around Belém in places such as Macapá (Amapá) further north or Santarém (Pará) up the Amazon River.

Through a geolinguistic lens, the evolution of the palatalization of implosive /s/ observed in Brazil seems to be a partial repetition of what started in southern Portugal in the 17th century, only about 200 years later and in a graded way. The Brazilian South is linguistically conservative with little to no *chiamento*, while the intermediate zone shows the beginning of the evolution, and minor zones have reached the end of the process with complete *chiamento*. Rio de Janeiro was never a unique or isolated case within Brazil, linguists only had rather little knowledge of Brazilian varieties before. Considering linguistic testimonies from the 19th and early 20th centuries as well as the situation in Belém do Pará, it becomes extremely unlikely that the short presence of the Portuguese Court exerted much influence on the formation of *chiamento* in Rio de Janeiro. Had this really been the case, today's pronunciation of *descer* and *nascimento* would have to include [-ʃs-] like in Portugal. This nexus, however, is absent in Rio de Janeiro.

5 Diachronic evolution leading to structural differences in EP and BP phonology

When comparing the sound systems of EP and BP (see chapter 12, Sound-related aspects of Brazilian Portuguese), it is important to note that EP has a well-defined standard; BP, on the contrary, depends largely on what Brazilian speakers tacitly approve of without necessarily practicing it themselves. For instance, there has been no general ruling on the realization of implosive /s/ since the *Primeiro Congresso Brasileiro da Língua Falada no Teatro* (1958) accepted *chiamento* (see Section 4.1) for theater productions in 1958.

In Recife, local educated speakers do not affricate /t/, /d/ in font of [i], although a majority of Brazilians do. University lecturers from São Paulo will not necessarily suppress the *r-caipira* (see Section 3.2). Brazilian speakers and authorities obviously do not feel the need for any sort of standardized pronunciation such as RP (received pronunciation) in British English. In contrast to regional influence, substandard pronunciations are avoided in formal situations. *Mulher* [mu'lɛ] ([l] for [ʎ] in this particular word) would still be considered quite colloquial, [mu'jɛ] and ['vɛju] (*velho*) are clearly marked. This also applies to rhotacisms (*claro* > ['kraru]).

The following sections summarize the diverging diachronic evolution of EP and BP, with differences marked in gray.

5.1 Vowels

Stressed oral vowels: There is a minor difference between stressed oral vowels in EP and BP which concerns the phonological status of /ɐ/ in EP (see Section 3.1). The opposition /a/ vs. /ɐ/ in EP dates from the 16th century and is hardly relevant within the system (distinction between present and simple perfect tenses as well as between *a* and *à*). In BP, [ɐ] is only an allophone. Consequently, the number of vowel phonemes is seven in BP versus eight in EP. In contact with a following nasal consonant, BP prefers the close-mid allophones [e] and [o] (BP *prêmio* [e] vs. EP *prémio* [ɛ] < Lat. *ae* → Port. [ɛ]), whereas EP maintains the difference between close-mid and open-mid vowels in this position (cf. EP *sono* [o] vs. EP *fome* [ɔ]).

Tab. 1: Stressed oral vowels BP—EP

BP			EP		
i		u	i		u
e		o	e		o
ɛ	[ɐ]	ɔ	ɛ	ɐ	ɔ
	a			a	

Pretonic vowels: In BP, pretonic vowels appear in two different sets of five units with either close-mid [i e a o u] or open-mid vowels [i ɛ a ɔ u], the open-mid pronunciation being more frequent in the Northeast. In EP, there are eight units ([i ɛ ə ɐ a ɔ o u]).

Post-tonic vowels: In post-tonic (non-final) position, BP once again displays five units in two sets ([i e a o u] and [i ɛ a ɔ u]) compared to four units in EP ([i ɐ ə u]). The commonly held view previously was that post-tonic vowels in BP did not comprise open-mid units, but the *Atlas Linguístico do Brasil* (AliB) shows that open-mid vowels also appear, though mainly in the Northeast. Post-tonic [ɔ] also exists in

Pará, Minas Gerais, Espírito Santo and locations in the South (cf. ALiB, cartas F02 V 1, F02 V 3).

Unstressed final vowels: The set of unstressed final vowels comprises three units both in BP ([-i -a -u]) and in EP ([-ə -ɐ -u]) if we leave aside an exception like *táxi*. The local appearance of final [-e], [-o] in rural areas of São Paulo and the South is not really significant, whereas the occasional raising of final /a/ > [-ɐ] in BP needs further investigation (see Section 3.1). The reason for the difference between EP and BP in unstressed vowels lies mainly in EP centralizing/raising them in the 18th century ([e] → [ə] / [ɨ], [a] → [ɐ], [o] → [u]). In this respect, the BP system proves to be more balanced than that of EP.

Tab. 2: Pretonic oral vowels BP—EP (*with regional variants)

BP			EP		
i		u	i		u
	e	o		ə	o
	*ɛ	*ɔ	ɛ	ɐ	ɔ
	a			a	

Tab. 3: Post-tonic oral vowels BP—EP (*with regional variants)

BP			EP		
i		u	i		u
	e	o		ə	
	*ɛ	*ɔ		ɐ	
	a				

Tab. 4: Unstressed final oral vowels BP—EP (*with variant)

BP			EP		
i		u	(i)		u
				ə	
	*ɐ			ɐ	
	a				

Oral diphthongs: BP has at least two additional oral diphthongs as a result of the vocalization of implosive /l/. The diphthongs concerned are [ɔu̯] (*sol*) and [uu̯] (*sul*). The third diphthong is [ou̯] (*bolso*) if we disregard the use of the archaic variant [ou̯] for

<ou> (*vou*) in BP and the regional presence of [oʊ̯] in northern Portugal. The substitution of [ɐi̯] for [ei̯] in EP is due to vowel opening in the 19th century. Regionally, we also find [ɐi̯] in BP when in contact with a nasal (e.g., *paina*).

Tab. 5: Oral falling diphthongs BP—EP (*with regional variants)

BP						EP				
iʊ̯				ui̯	uʊ̯	iʊ̯				ui̯
eʊ̯	ei̯		oi̯	oʊ̯		eʊ̯	*ei̯		oi̯	*oʊ̯
ɛʊ̯	ɛi̯	*ɐi̯	ɔi̯	ɔʊ̯		ɛʊ̯	*ɛi̯	ɐi̯	ɔi̯	
	ai̯	aʊ̯					ai̯	aʊ̯		

Tab. 6: Nasal vowels and diphthongs BP—EP

BP					EP				
ĩ			ũĩ̯	ũ	ĩ			ũĩ̯	ũ
ẽ	ẽĩ̯		õĩ̯	õ	ẽ			õĩ̯	õ
		ɐ̃					ɐ̃		
	ɐ̃ĩ̯	ɐ̃ũ̯				ɐ̃ĩ̯	ɐ̃ũ̯		

Nasal vowels and diphthongs: In the area of nasal vowels and diphthongs, the main difference between EP and BP is the substitution of [ẽĩ̯] for [ẽĩ̯] in EP. This is also due to vowel opening in the 19th century. Previously, [ẽĩ̯] in EP only existed in the word *mãe* [mẽĩ̯]. Apart from that, BP maintains a regionally varied kind of heterosyllabic nasalization triggered by a nasal consonant of the subsequent syllable (BP *ca̲ma* ['kɐ̃.ma] vs. EP ['kɐ.mɐ]).

5.2 Consonants

Brazilian voiced plosives ([b d g]) are plosives in all positions, whereas they formed fricative allophones in EP ([β ð ɣ]), especially /d/, during the 16th century. Two archaic features in BP consonantism are the affricate [tʃ], which has also been maintained in northern Portugal, and velarized implosive [ɬ] encountered in standard EP. Both only appear regionally.

As for palatalization and the evolution of rhotics, Brazilian consonantism is innovative. The affrication of /t/, /d/ in front of [i] has produced the two allophones [tʃ], [dʒ] which are a major feature of BP. The palatalization of implosive /s/ in parts of the country is also an independent innovation. Velarization and glottalization of /r̄/ are

represented by the phoneme /r̄/ [h] with quite a number of free regional variants in southern Brazil ([r̄ ɣ x r]). Implosive /r/ has mainly become [x] or [h], the retroflex post-alveolar approximant *r-caipira* [ɻ] being a free regional variant.

Tab. 7: BP consonants with [allophones] and some *regional variants

	bilabial	labiodental	dental	alveolar	prepalatal	palatal	velar	uvular	glottal
plosive	p		t				k		
(voiced)	b		d				g		
fricative		f		s	ʃ		[x]		[h]
(voiced)		v		z	ʒ				
affricate					tʃ				
(voiced)					dʒ				
nasal	m			n		ɲ	[ŋ]		
lateral				l		ʎ	*ɬ		
vibrant (multiple)				r					
				*r̄					
approx.				*ɻ		j	w		

Tab. 8: EP consonants with [allophones] and some *regional variants

	bilabial	labiodental	dental	alveolar		prepalatal	palatal	velar	uvular	glottal
plosive	p		t					k		
(voiced)	b		d					g		
fricative		f		s	*ʂ	ʃ			[ʁ]	
(voiced)	[β]	v	[ð]	z	*ʐ	ʒ		[ɣ]		
affricate						*tʃ				
(voiced)										
nasal	m			n			ɲ	[ŋ]		
lateral				l			ʎ	[ɬ]		
vibrant (multiple)				r						
				*r̄						
approx.							j	w		

5.2 Suprasegmental aspects

There is no documentation for the evolution of intonation in BP. The only comparative remark found in older literature is the one by Alexander Caldcleugh in 1825:

> The Portuguese spoken by the Brazilians is easily distinguishable from that used by the natives of the mother country. The mode of speaking is much slower, a peculiarity to be observed in all colonies [...] The pronunciation of the Brazilians is not so nasal [...], and on the whole it is a more agreeable language than in the mouth of a native [of the mother country] (Caldcleugh 1825, I, 65–66).

This seems to correspond to Fernão de Oliveira's observation about Portuguese diction in 1536: "e outras nações cortam vozes apressando-se mais em seu falar, mas nós falamos com grande repouso como homens assentados" (Oliveira 1975, 39) 'and other nations shorten words in order to accelerate their speech, whereas we talk taking our time as determined men do'.

As for syllable structure, spoken BP displays a clear tendency, already documented in the 18th century, to break up consonant clusters (plosive + fricative/nasal), thereby attaining a higher number of basic CV-syllables. In informal speech, the procedure extends to metathesis. EP has maintained the clusters in learned words and weakened most unstressed vowels in the 18th century. It was the reduction of [e], [i] > [ə] in particular that weakened EP syllable nuclei. This vowel loss, especially prominent in the speech of Lisbon, is prone to then diminish the number of syllables and increases consonant clusters (cf. *Cristina* ['kʃtinɐ]). The loss also affects final syllables (*arte* EP ['art], BP ['ax.tʃi]). For this reason, the phonetics of EP are often considered somewhat slurred, while BP gives the impression of clear vocalism in easily discernable syllables.

6 Conclusion

Brazilian vocalism mainly preserves the historic quality of Portuguese vowels up to the 18th century when EP started reducing unstressed units as well as centralizing [e] +palatal and [ei̯], [ẽĩ̯] in standard language in the 19th century. Pretonic variation in BP mirrors the lack of standardization in the past, whereas new diphthongs originated due to the vocalization of implosive /l/. The introduction of epenthetic vowels and substandard metathesis show the predominance of spoken language in the development of BP. Apart from preserving original implosive alveolar /s/ [s], especially in southern parts of the country, Brazilian consonantism has been mainly innovative from the 18th and 19th centuries on as reflected in the widespread affrication of /t/, /d/ before [i], the vocalization of implosive /l/ and the velarization or glottalization of rhotics. Partial or general palatalization of implosive /s/ [ʃ] and a certain tendency to close final /a/ [-ɐ] in present-day BP repeat previous developments in EP.

7 References

ALiB: Suzana Alice Marcelino da Silva Cardoso et al., *Atlas Linguístico do Brasil. Cartas linguísticas 1*, II, Londrina, Editora da Universidade Estadual de Londrina, 2014.

Azevedo, Milton M. (2005), *Portuguese: A Linguistic Introduction*, Cambridge, Cambridge Univ. Press.

Balbi, Adrien (1826), *Introduction à l'atlas ethnographique du globe, contenant un discours sur l'utilité et l'importance de l'étude des langues appliquée à plusieurs branches des connaissances humaines* [...] I, Paris, Rey et Gravier.

Barbosa, Domingos Caldas (1826), *Viola de Lereno: collecção das suas cantigas, offerecidas aos seus amigos*, II, Lisboa, Lacerdina.

Barros, João de (1540), *Grammatica da lingua Portuguesa*, Olysippone, Apud Lodouicum Rotorigiũ Typographum.

Caldcleugh, Alexander (1825), *Travels in South America, During the Years 1819–20–21; Containing an Account of the Present State of Brazil, Buenos Ayres, and Chile*, 2 vol., London, Murray.

Carmelo, Frei Luís do Monte (1767), *Compendio de orthographia*, [...], Lisboa, Antonio Rodrigues Galhardo.

Coutinho, Jozé J. da Cunha de Azeredo (1798a), *Estatutos do Seminario Episcopal de N. Senhora da Grasa da Cidade de Olinda de Parnambuco* [sic], Lisboa, Na Typografia da Acad. R. das Ciencias.

Coutinho, Jozé J. da Cunha de Azeredo (1798b), *Estatutos do Recolhimento de N. Senhora da Gloria do lugar da Boa-Vista de Parnambuco* [sic], Lisboa, Typografia da Acad. R. das Siencias.

Entwistle, William J. ($^{4/1982}$ 1982, ^{1}1973 [1936]), *Las lenguas de España: castellano, catalán, vasco y gallego-portugués*, Madrid, Istmo.

Giangola, James P. (2001), *The Pronunciation of Brazilian Portuguese*, München, Lincom.

Kröll, Heinz (1994), *Portugiesisch: Dialektale und regionale Varianten in Portugal. Dialectos e variedades regionais em Portugal*, in: Günter Holtus/Michael Metzeltin/Christian Schmitt (ed.), *Lexikon der Romanistischen Linguistik* (LRL), vol. VI/2, Tübingen, Niemeyer, 545–559.

Leão, Duarte Nunes de (1576), *Ortografia e origem da língua portuguesa. Introdução, notas e leitura de Maria Leonor Carvalhão Buescu*, Lisboa, Nacional.

Leite, Yonne/Callou, Dinah (2002), *Como falam os brasileiros*, Rio de Janeiro, Zahar.

Lipski, John M. (1975), *External History and Linguistic Change: Brazilian Portuguese -s*, Luso-Brazilian Review 12, 213–224.

Lipski, John M. (1976), *Final s in Rio de Janeiro: Innovation or Imitation?*, Hispanic Review 44, 357–370.

Maia, Clarinda de Azevedo (1986), *História do galego-português. Estado linguístico da Galiza e do Noroeste de Portugal desde o século XIII ao século XVI (Com referência à situação do galego moderno)*, Coimbra, INIC.

Maia, Manuel Rodrigues (ca. 1800), *O periquito ao ar, ou O velho uzurário*. Farça que se reppresentou no Theatro Nacional do Salitre com geral acceitação, composta pelo profeçor Manoel Rodrigues Maia, ampliada por por Alexandre Jozé Victor da Costa Sequeira, copiada aos 19 de Janeiro de 1818.

Markgraf, Georg/Piso, Willem (1848), *Historia naturalis Brasiliae*, [...] Lugdun. Batavorum, Hackium—Amstelodami, Elzevirium.

Moraes, João A./Wetzels, W. Leo (1992), *Sobre a duração dos segmentos vocálicos nasais e nasalizados em português: um exercício de fonologia experimental*, Cadernos de estudos linguísticos 23, 153–166.

Mott, Luiz R. B. (1979), *Uma escrava no Piauí escreve uma carta [...],* Mensário do Arquivo Nacional 5, 7–10.

Nascentes, Antenor (1921), *Variante carioca de um subdialecto brasileiro*, Revista do Brasil 18, 312–321.

Nascentes, Antenor (^{2}1953, ^{1}1922), *O linguajar carioca*, Rio de Janeiro, Simões.

Noll, Volker (1996), *Rio de J*aneiro *und das "chiamento" im brasilianischen Portugiesisch*, Iberoromania 43, 1–9.
Noll, Volker (2008), *O português brasileiro. Formação e contrastes*, São Paulo, Globo.
Oliveira, Fernão de (1975 [1536]), *A gramática da linguagem portuguesa. Introdução, leitura actualizada e notas por Maria L. Carvalhão Buescu*, Lisboa, Imprensa Nacional.
Oliveira, Klebson (2003), *Textos escritos por africanos e afro-descendentes na Bahia do século XIX: fontes do nosso latim vulgar?* 3 vol., Salvador [diss. de mestrado].
Pessoa, Marlos (1994), *"A lingoagem bordalenga de muita gente": o conteúdo lingüístico de importante fonte para o conhecimento do português brasileiro do século XIX*, Lusorama 25, 70–80.
Révah, Israel S. (1958), *L'évolution de la prononciation au Portugal et au Brésil du XVIe siècle à nos jours, in: Anais do Primeiro Congresso Brasileiro de Língua Falada no Teatro. Bahia—1956*, Rio de Janeiro, Ministério da Educação e Cultura, 387–402.
Rhys, John D. (1992 [1569]), *De Italica pronvnciatione, & orthographia*, in: Nicoletta Maraschio (ed.), *Trattati di fonetica del Cinquecento*. Firenze, Crusca, 91–193.
Salvador, Frei Vicente do ([7]1982 [1627]), *História do Brasil, 1500–1627*, São Paulo, Itatiaia.
Seraine, Florival (1949), *Aspectos históricos da língua nacional no Ceará*, Revista do Instituto do Ceará 63, 49–74.
Silva, José J. Paranhos da (1879), *O idioma do hodierno Portugal comparado com o do Brazil. Por um brazileiro*, Rio de Janeiro, Winter.
Silva, José J. Paranhos da (1880), *Sistema de ortographia brazileira*, Rio de Janeiro, Winter.
Silva Neto, Serafim da ([5]1986, [1]1950), *Introdução ao estudo da língua portuguesa no Brasil*, Rio de Janeiro, Presença.
Stavrou, Christopher (1947), *Brazilian-Portuguese Pronunciation Including Word List With Indicated Pronunciation*, Philadelphia, McKay.
Teyssier, Paul (1966), *La prononciation des voyelles portugaises au XVIème siècle d'après le système orthographique de João de Barros*, Annali dell'Istituto Orientale di Napoli (Sezione Romanza) 8, 127–198.
Teyssier, Paul ([2]1984, [1]1975), *História da língua portuguesa*, Lisboa, Sá da Costa.
Verney, Luís A. (1746), *Verdadeiro metodo de estudar, para ser util à Republica, e à Igreja,* 2 vol., Valensa, Antonio Balle. [written by 1736].
Viana, Aniceto R. Gonçalves (1883), *Essai de phonétique et de phonologie de la langue portugaise d'après le dialecte actuel de Lisbonne*, Romania 12, 29–95.

Mário Eduardo Viaro

5 Historical morphology

Abstract: The development of synchronic studies in the 20th century led to the emergence of multiple theoretical approaches to morphology, including a better understanding of diachronic data from old historical grammars. Given that morphology tends to be rather conservative in comparison to the lexicon, differences between Brazilian and European Portuguese must be understood as a continuum between similarity and difference, and morphology tends to maintain more similarities than the lexicon. Apart from the standard language, relevant data can also be found in colloquial BP. In addition to traditional descriptions of changes in the inflectional system and differences in the pronominal system and its morphosyntactic behavior, there is a lot of room for the theoretical integration of proposals from stylistics, semantics and pragmatics in describing BP innovations, especially in the spoken language.

Keywords: morphological paradigms, extended semantics, diachronic morphology, neology, sign reference

1 Introduction

The term "morphology" was adopted by Comparative Linguistics in the second half of the 19th century. Since then, inflectional phenomena have been examined more extensively than their derivational counterparts in diachronic studies, and formal views are more common than studies on content. Classic studies on Portuguese historical linguistics, such as Cornu (1888), Nunes ([1]1906; [1]1919), Huber (1933), and Williams (1938), do not detail the semantic features of each affix, even though descriptions of formal changes in inflectional morphology are abundant. Relevant exceptions can be found in Leoni (1858) and Said Ali (1921; 1923). From the second half of 20th century onwards, morphology studies have largely focused on synchrony. Exceptions to this almost exclusive point of view can be found, such as in Rio-Torto (1993; 1998). It is not unusual for synchronic and diachronic problems to merge, due to the absence of consensual presuppositions in morphology: for instance, are derivations a synchronic process or a diachronic one? Since the synchronic perspective on morphological phenomena also implies a temporal dimension (when *primitive* and *derived* forms are described), those problems are still not coherently answered by the current theorization of morphology.

Insufficient studies on derivational historical morphology are not the only problem. Another one is that BP is normally studied outside of the comparative perspective. Recent works like Rocha (1999), Ilari (2014; 2015) and Rodrigues/Alves (2015) offer a lot of information based on use of the so-called Brazilian *norma culta* 'standard

https://doi.org/10.1515/9783110405958-006

language', but do not emphasize notable differences between Brazilian and European Portuguese, nor do they take a historical approach in order to understand morphological productivity: some examples that are mentioned were formed in past synchronies and are merged within lexical paradigms in the most recent synchrony. This, however, is a current trend. When certain salient phenomena in the productivity of affixes are described, it is often the case that unstable and less representative forms, present only in the *parole*, are mixed with stable ones, making it difficult to achieve the desired comparisons of BP structures with European Portuguese (EP), Angolan Portuguese and so on.

Certain other particular difficulties, not identical to those in syntax or phonology, arise from the existing studies. For instance, a large number of morphological exceptions can frequently be found in Indo-European languages, and these have always been a challenge for scientific and theoretical generalizations in linguistics. Thus, we know that BP x-*al* is a suffix which denotates "a place where x trees are planted" and that its variant *-zal* is used in a morphophonological context in which *x* has the last syllable stressed (as *laranja* 'orange' → *laranjal* 'a plantation of orange trees' and *café* 'coffee' → *cafezal* 'a plantation of coffee trees'). How is a formation like *mangue* 'mangrove' → *manguezal* 'mangrove vegetation' possible? *Mangue* does not have its syllable stressed, nor is *manguezal* exactly a plantation, although the meaning of the word describes a place or a collection of some special trees. Is the suffix *-zal* a homonym or could some semantic similarity allow us to say that it has the same, perhaps less prototypical, meaning? Different solutions for the same problem can be proposed, since the theoretical presuppositions on which they are based are not always firmly explicit.

The acceptance of abstract units such as "morphemes", which are ideal representations of more concrete forms, or "morphs", was a typical structuralist solution to synchronic problems (Coelho/Lopes 2015), but again exceptions show us that the establishment of an exhaustive list of morphemes is extremely difficult (Heckler/Back/Massing 1984; 1994): why does *ladrão* 'thief' have a feminine form *ladra* if the usual corresponding morphemes of feminine gender for words ending in *-ão* are *-ã*, *-oa* or the more productive *-ona*? In fact, in BP, the word *ladra* is sometimes replaced by the variant *ladrona*, in order to overcome the irregular situation (not only of suffixation, but also of stress changing) of a one-item paradigm. But without etymological studies it is hard to say whether that replacement was made first in the system of BP or if it was inherited from common Portuguese.

In Generative Linguistics, the problem of finding a solution to irregularity can be avoided by interpreting exceptions as belonging to a "lexicon" (and hence they should be studied within lexicology, not morphology). Some authors understand morphological exceptions as "degenerate versions" of the rules for forming new words and elect one characteristic, i.e., productivity, as a milestone for the determination of the object of morphology *versus* lexicology (Aronoff 1976, 34). Regularity might be a reality in highly agglutinant languages such as Turkish or for languages with small numbers of inflections and with few derivative suffixes, like English, but not for Rus-

sian or Portuguese. In fact, irregularities are also not a major concern for other linguistic disciplines, like lexicology or syntax. The same could be said of the order and internal agreement between irregular morphs. The only branch of linguistics which could answer these questions is morphology. And historical morphology still offers the best way to explain the existence of reminiscent irregularity and unproductive morphemes.

Since the final decades of 20th century, some morphologists in Brazil have developed a special metalanguage as a response to the understanding of morphology as a subset of phonology or syntax. In fact, morphophonology and morphosyntax would be possible only if the terminology used by morphologists were clearly defined. There have been few systematic or exhaustive studies on the historical morphology of derivational Portuguese elements so far. This is because etymological research in Portuguese was less developed than in other major European languages, and it is not possible to reach conclusions in diachronic studies without good etyma or etymological data. However, there is some pioneering work, such as Cunha (1982; 2006), who looked at the dating of the oldest attestations in Portuguese (the so-called *terminus a quo*). His research was then adopted by Houaiss/Villar (2001). Studies on historical morphology have been developed since then, but, unlike the case of Spanish (Pharies 2002), there is still no etymologically well-based and comprehensive work on Portuguese morphological phenomena such as affixation.

Studies about Portuguese historical morphology (and not only BP) were developed by the following Brazilian research groups: (1) GMHP at the University of São Paulo/Brazil:[1] Areán-García (2007; 2011; 2013), Lacotiz (2007; 2013), Viaro (2007; 2011c; 2013a), Freitas (2008; 2013; 2014; 2018), Gianastacio (2009; 2013; 2015), Gonçalves (2009; 2013), Santos (2010; 2016), Rio-Torto (2013), Becker (2013), Takahashi (2013), Oliveira (2014), Santana (2017); and (2) PROHPOR, at the University of Salvador/Brazil,[2] such as Coelho (2000; 2005), Campos (2004), Santana (2008), Santos (2009), Lopes (2013; 2018) and Simões Neto (2016; 2019). For exclusively BP phenomena, however, synchronic studies are most commonly available.

The remainder of this chapter is organized as follows: innovations in the morphological paradigms of BP will be discussed in section 2, and the inheritance of word formation and word classes from EP is detailed in sections 3 and 4. Prior to setting out conclusions, the emergence of new meaningful morphological elements from meaningless phonological segments will be illustrated in section 5.

1 www.usp.br/gmhp, last accessed 16.02.2022.

2 https://www.prohpor.org/morfologia-lexico-historicos, last accessed 16.02.2022.

2 Innovation in morphological paradigms

Word classes exhibit distinct diachronic behavior, and the consequences of this can be seen in BP. Closed word classes such as articles, numerals, prepositions and inflectional endings generally change very slowly. In BP, there were only sporadic changes in the inherited inflections of gender and number. Some variation in the morphosyntactic phenomenon of number agreement does occur, however: in spoken BP it is common not to use redundant number agreement in the subordinated elements of a nominal syntagma (as *os cachorro* instead of *os cachorros* 'the dogs'), in the relation subject-verb (*os cachorro late* instead of *os cachorros latem* 'the dogs bark') or in the predicative (*os carro dele são tudo importado*, instead of *os carros dele são todos importados* 'his cars are all imported'), but absence of agreement in gender is substandard and regional (see ↗ 6 Historical syntax). Amaral (1920) gives examples: *essas coisarada bonito*, instead of *essas coisas bonitas* 'these beautiful things', *as criança távum quéto* instead of *as crianças estavam quietas* 'the kids were quiet'. More recent examples can be found in Lucchesi (2009).

Major differences with respect to EP are seen in pronouns: *tu* 'you' is used in many Brazilian regions together with standard BP *você* 'you' (*vosmecê*, *mecê*, *vassuncê* are considered old-fashioned forms today, but *ocê* and *cê* are very common in speech), and *vós* is only used in religious language. As a direct complement, *te* 'you' is common and mixed with *você* in subject and post-prepositional uses. In BP, the post-prepositional pronoun *si* as synonymous with 'you' is never used, as in EP, except as a reflexive meaning 'yourself', 'himself', 'herself', 'themselves'. The pronouns *ele*, *ela*, *eles*, *elas* are used instead of *o* 'him', *a* 'her', *os* 'them (masc.)' and *as* 'them (fem.)' in colloquial language. This is a very important difference between colloquial BP and EP, but it is also a question of comparative frequency of use in written language. Historically, we cannot claim any innovations in the forms of the pronoun system of BP, unless some yet more colloquial uses are considered: *o papai aqui* lit. 'the dad here' can be used as colloquially synonymous to *eu* 'I'; *os caras* lit. 'the fellows' ou *os nego* 'the black guys' is sometimes heard as colloquially synonymous to *eles* 'they'. Further examples of common evaluative uses of nouns instead of pronouns are: *o filho da mãe é inteligente* 'he is intelligent' (literally, 'the son of bitch'), *o bicho ficou uma fera* 'he became very angry' (literally: 'the beast'), *o povo arrebentava de rir* 'they burst out laughing' (literally, 'the folk'), *neguinho não entende* 'he/they don't understand' (literally: 'little black'), *a pessoa adoece*, *se faz isso* 'you sicken, if you do it' (literally, 'the person'). Those forms are similar to the grammaticalized emergence of *a gente* 'we' in the 19th century (Lopes 1998; 2003; see also ↗ 6 Historical syntax) or *vossa mercê* 'you' (literally: 'your mercy') and other honorific expressions in the early 16th century. Such cases are normally interpreted as examples of stylistic expressiveness and not as morphological innovations, but some of them are turned into crystalized forms and are in fact strongly grammaticalized in speech by phonetic erosion.

The so-called open word classes are even more innovative: by means of nouns, adjectives, and verbs, the lexicon of BP is clearly distinguished from EP. A large proportion of those lexical items are peculiar derivate forms. The existence of some peculiar affixes or typically Brazilian uses of affixes will be discussed in the following sections.

Closed and open classes are normally related to traditional word classes, but this is not absolutely the case: for instance, personal pronouns are more conservative than indefinite pronouns, which really form an open subclass. Very few Latin indefinite pronouns survived in all Romance languages. It is not possible to reconstruct Indo-European indefinite pronouns as we can do with personal pronouns. In Latin *nihil* 'nothing' comes from an expression which means 'not even the eye of a bean', but NIHIL is not the etymon of Portuguese and Spanish *nada*, which is (REM) NATAM 'something born' (which is also different from French *rien*, from Italian *nulla* and from Romanian *nimică*).

Changes in morphological classes cannot be considered without acknowledging differences in the potential for expressiveness among the subsets of each class. A negative adverb like *não* 'no, not' is conservative in that it stems from Latin NON, but a peculiar morphosyntactic variation should be noted. In order to say 'I do not know', besides normative *eu não sei* 'I not know' plenty of other utterances can be heard, such as *não sei não* (literally 'I do not know, no'), *sei não* (literally 'I know not'), *sei lá* (literally 'I know there'), *sei lá eu* (literally 'I know there, me'), *eu é que sei?* (literally 'is it me who knows?'). Some of those forms are not exclusive in BP, yet to state which are inherited from EP (and which are still used in Portugal as normal speech or as dialectal expression) is difficult due to a lack of comparative research.

Emphatic negation, however, creates new forms, and from a diachronic perspective we can say that new negative adverbs can easily emerge, acting as an open subclass. In fact, the potential of expressiveness of negation sometimes becomes more important in a sentence than its logical function. This is particularly noticeable in the case of negative indefinite pronouns, which commonly mix negation with expressive exaggeration. For instance, in certain sentences pronouns like *nenhum* 'none' or *nada* 'nothing' really mean 'a few' or 'very few' and there are lots of innovative forms in BP speech: instead of saying *não entendi nada* 'I did not understand anything', other variants are typically heard: *não entendi coisa nenhuma* (literally, 'no thing'), *não entendi nadica de nada* (literally, 'the little nothing from nothing'), *não entendi bosta nenhuma* (literally, 'no shit'), *não entendi porra nenhuma* (literally 'no cum'), *não entendi pissirica* (literally 'female masturbation') or even expressive words without meaning in recent synchrony: *não entendi patavina*, *não entendi bulhunfas*. As in the English phrase 'I did not understand a damn thing' and in other languages, the use of obscene words is also a means of adding intensification to an expression of nullity.

Further examples: *um negócio assim* (literally 'such a business') is a colloquially crystalized option instead of the formal *algo assim* 'something like that'. Stylistically reinforced forms are frequently used in speech and are themselves a source to new elements in those word classes: *ele conhece tudo quanto é coisa* 'he knows everything

at all' (literally 'everything which is a thing'); *ele conhece coisa pra caramba* 'he knows very much' (literally 'things for darn/damn'); *ele sabe de um tudo* 'he knows everything' (literally 'a part of everything').

The same can be said of conjunctions, which are also rather innovative, even though some conjunctions expressing logical functions are more resistant to language change (such as Portuguese *e* < ET 'and', *ou* < AUT 'or', *se* < SI 'if'; for the case of *caso* 'case' > 'if' see Gerards/Kabatek 2018). Concessive conjunctions are more expressive, because they form an open subclass. In fact, Romance languages don't preserve forms inherited directly from Latin (as ETSI, QUAMVIS, QUAMQUAM, LICET); in Portuguese we find for instance *embora*, *conquanto*, *ainda que*, *mesmo que*, *se bem que*, *apesar de que* and so on. The same happened with adversative conjunctions, which are semantically related to concessives: one of the most used Brazilian adversative conjunctions is *só que* 'but' (literally, 'only that'), as common as the older *mas* 'but', which replaced Latin conjunctions like SED, TAMEN, AT, AST, VERO, VERUM and Old Portuguese *pero*, *empero*, *emperol*, *pero que*, *macar*, *siquer* (Nunes [1]1919; Longhin 2003). In BP, the old adversative form *mais* is also ubiquitous.

A reasonable explanation for the differences between conservative and innovative word (sub)classes is not expected to be found in a supposed 'essence' of the word classes, but instead in their potential for being 'expressive'. If, on the one hand, classes which are particularly useful for sheer description tend to be conservative, on the other hand, where judgement is concerned, a second level of semantic description arises together with innovations. Old forms are abandoned, and new ones, with more potential of expression, emerge in order to cause surprise or to be convincing. Again, it is a hard task to determine which of those forms are inherited from EP (and which are nowadays unusual) and which were born in BP, be they used exclusively in Brazil or adopted also by speakers in Portugal (by means of recent TV series or internet media, for instance). Only more exhaustive historical studies can shed light on these hitherto almost unresearched topics.

3 Word formation inherited by Brazilian Portuguese

Research into affixation is shared by morphology and lexicology. This leads us to the question of lexical segmentation. A segmentation is possible only by using synchronic knowledge (and this is what speakers nowadays do with the most-used words), because linguistics in a historical perspective does not primarily deal with communication. A verb like *comíamos* can be segmented as *com-í-a-mos* only when a synchronic morpheme commutation is applied, based on current language: *com-* would be the stem, *-í-* is the thematic vowel (instead of *-a-* as in the so-called "first conjugation"), *-a-* is an allomorph of the corresponding first conjugation morpheme *-va-*, and *-mos* is the morpheme for first person plural (agglutinated in all tenses of all verbs). If we consider the etymon of *comíamos* (Latin COMEDEBAMUS), it turns out that *com-* was diachro-

nically not a verbal stem but rather a prefix, and the original stem *-ed-* merged into the present thematic vowel. Despite this, traces of the original root of the Latin participle COMESTUS 'eaten' can be detected in the old Portuguese participle *comesto* 'eaten' (analogically changed to modern Portuguese *comido*), but its stem is residually present in modern Portuguese *comestível* 'edible' (although the infinitive root is found in the partial synonym *comível*). We know that synchronic phenomena are not altogether timeless. Strictly speaking, a synchrony is an arbitrary and methodologically chosen time span and we can speak about diachrony only when two or more synchronies are involved in a linguistic explanation. A segmentation may not be anachronistic, but anchored in a certain (past or present) synchrony, and the only way to avoid anachronism is the accurate description of each past synchrony which are used in the diachronic explanation of the lexical item, i.e., in its etymology. For accuracy, in analyses and formulae concerning historical morphology, synchronic phenomena as derivation and borrowing must be distinguished from diachronic phenomena: derivation will always be noted by the symbol → and a new symbol for loanwords ▸ is proposed, in order to differentiate both situations from the traditional symbol >, which would better be restricted to diachronic phonetic or phonological change (Viaro 2010a; [1]2011a; see also Viaro/Bizzocchi 2016, for the urgent need for a new and more accurate notation and system of symbols for etymological discussions).

In a derivation, a primitive form generates a new one. Although time is involved, both lexical items coexist after derivation in the same synchrony. This does not necessarily happen when two sequential synchronies are analyzed, that is, in diachrony (Viaro 2012; Viaro 2013b). Etymology searches for both relationships of cause and effect between lexical items within a synchrony or in subsequent synchronies. Sometimes, in order to affirm an etymological relation between two elements of different synchronies, the existence of a word or its existence over time must be proved, and when this is not possible, reconstruction is needed. A word always has a *terminus a quo* (the first date by which one can prove its existence), but its *terminus ad quem* (its last recorded use) is not so simple to determine, because words are often resurrected through writing. Etymology works with both heritage (diachrony) and communication (synchrony): BP and EP share a large number of lexical items inherited from pre-16th century Portuguese, which was also obviously a set of variant systems. Besides, two lexical items in a synchrony can have the same ending but not necessarily the same etymological trajectory of the suffix. For instance, Portuguese words ending in *-eiro* (Viaro 2007) can be:

(a) inherited words with an ending corresponding to a Latin word ended in *-arium*, e.g. Latin COQUINARIUM > Portuguese *cozinheiro* 'cook';
(b) an adaptation of the ending of a loanword, e.g., ancient French ENGEIGNEUR (modern French *ingénieur*) ▸ Portuguese *engenheiro* 'engineer';
(c) the synchronic result of a productive meaning of *-eiro*, e.g., Portuguese ABACATE → Portuguese *abacateiro* 'avocado tree';
(d) cases of analogy or coincidence, e.g., Tupi MAKAXÉRA ▸ Portuguese *macaxeira* 'manioc shrub'.

These examples exhibit a frequent phenomenon, that is, the formal convergence of endings into one suffix. When the Portuguese suffix *-agem* is compared to its Romance cognates, some differences are noticed: (a) by means of that suffix, feminine nouns are formed (instead of normal masculine Romance nouns: compare Portuguese *a viagem* 'the voyage' with Spanish *el viaje*, French *le voyage*, Italian *il viaggio*); (b) there is an innovative nasal ending that does not exist in other Romance languages. A possible explanation for such peculiarities of modern Portuguese *-agem* (fem.) is a merging of neuter -ATICUM words > old Portuguese *-age* (masc.) with feminine plant names ending in -AGINEM > old Portuguese *-agem* (fem.), such as Latin PLANTAGINEM > *tanchagem* 'broadleaf plantain' (Gonçalves 2009).

Divergence is another frequent diachronic phenomenon: Latin -ARIUM developed two different Portuguese suffixes: *-eiro* and *-ário*. From a diachronic point of view, they are the same suffix. In fact, semantic parallels are found in divergent forms, and sometimes we need to consider them together to show parallel developments of meaning (Viaro 2007). In historical morphology, we must expect that polysemy and homonymy can live together. The Latin suffix *-arium* was inherited from the Italic languages, but there was also an ending *-arium* which is an adaptation of Greek words ending in -άριον (γλωσσάριον ▸ *glossarium* ≈ *glosarium*, λογάριον ▸ *logarium* ≈ *logarion*, βωλάριον ▸ *bolarium* ≈ *volarium*, ὠδάριον ▸ *odarium*, σπογγάριον ▸ *spongarium*). This ending was reinterpreted as a new meaning of *-arium*, normally 'little X' or 'a set of Xs' which is detected in the history of the Portuguese language (Viaro 2010b; The symbol ≈ indicates two synchronic lexical variants of the same word).

Another problem related to diachronic development is meaning. In case (d) above, *macaxeira* has a false suffix when it is considered diachronically, notwithstanding the fact that it is possible to extract a functional suffix by means of segmentation from a synchronic perspective. If the speaker considers that the suffix of *macaxeira* is the same as other plant names like *roseira* 'rosebush' or *mangueira* 'mango tree', the stem *macax-* will result in being an opaque morpheme, yet associated to the content of the lexical item, i.e., *mandioca* 'manioc'. This is also possible in *chiqueiro* 'pigpen', whose suffix seems to have the same meaning as the one in *galinheiro* 'hen house'. What is different is that the stem *galinh-* easily evokes the item lexical *galinha* 'hen', but *chiqu-* does not evoke the word *porco* 'pig', although it is not wrong to affirm that *chiqu-* is associated with *porco*. This associative meaning is obtained only by means of a subtraction of lexical meaning and the meaning of the suffix. The word *galinheiro* is compositional, because it is an example of an x-*eiro* structure (x is here the stem of the lexical item), where *-eiro* means 'place where x are kept', like others. In order to paraphrase words like *chiqueiro*, since they were compositional, we need the meaning of *-eiro* to be productive, and this is only possible synchronically. Nevertheless, we can call this *associative compositionality*. Conversely, a word like *chuveiro* 'shower' has the meaning of its stem directly attributed to *chuva* 'rain', but its *-eiro* suffix is opaque; in fact only *ad hoc* paraphrases would be possible and they would be uniquely used in this lexical item. The reason of such a situation is historical: the an-

cient meaning of *chuveiro* (still found in dialectal variants of BP) is 'heavy rain', and thus it is also found in *nevoeiro* 'heavy mist' and *poeira*, originally 'strong dust', nowadays 'dust'. Yet such a compositional interpretation of *-eiro* in *chuveiro* as 'large amount of X' is possible only in past synchronies or in local variants of BP. A word like *cadeira* 'chair' is not compositional, either by means of *ad hoc* attribution of meaning to a supposed *cad-* stem or a supposed *-eira* suffix or through associative attribution. In fact, *cadeira* has nothing to do with the original -ARIUM: the etymon of this word is the Latin *cathedra*, and its origin is Greek καθέδρα. Differently from *macaxeira*, *chiqueiro* and *chuveiro*, a supposed suffix of *cadeira* is not associated with anything, and only in this case can we call it a pseudosuffix in historical morphology.

The word *barbeiro* 'barber' has a stem *barb-* 'beard' (direct association) to which the suffix *-eiro* is aggregated with the specific meaning 'person who (v) x'. The meaning of v is pragmatically obtained by the same kind of association seen in previous examples: if it is interpreted as 'to shave', we correctly obtain the compositional meaning of the lexical item *barbeiro* as 'person who shaves beards'.

A compositional interpretation is supposed to be the original meaning of the word. Currently, however, the word *barbeiro* is sometimes used in BP for a very known species of insect (*Triatoma infestans*, order Hemiptera, family Reduviidae) or for a bad driver. These non-compositional meanings are not an object of study of morphology, but of lexicology. The classical distinction between homonymy and polysemy is used in the lexicographical task of building dictionaries. From this point of view, two identical forms (marked with the symbol ≡ as in Viaro [1]2011a) with different meanings are homonyms only if their etyma are different. We speak of polysemy, *stricto sensu*, only if two items have the same origin. A word like *barbeiro* is an example of polysemy. Its original meaning is 'barber' (*barbeiro*[1]). The meaning 'a specific insect' (*barbeiro*[2]) is a derived one. Nevertheless, both have the same origin (French *barbier*). The formula French BARBIER > Portuguese *barbeiro*[1] ≡ *barbeiro*[2] ← Portuguese BARBEIRO[1] < French BARBIER seems redundant. However, if it is rewritten (less preferably) either as French BARBIER > Portuguese BARBEIRO[1] → Portuguese *barbeiro*[2] or as French BARBIER > Portuguese *barbeiro* 'barber' >> 'insect' (the symbol >> is used for semantic change, cf. Viaro [1]2011a), it is not clear if *barbeiro*[1] and *barbeiro*[2] are still coexistent in the same synchrony. We can see that the etyma of both forms are not the same (the etymon of *barbeiro*[1] is French *barber* and that of *barbeiro*[2] is *barbeiro*[1]), but its same remote origin allows us to classify the two terms as a case of polysemy.

It is not uncommon for a meaning of a suffix to have its etymon in a non-compositional meaning of a word by means of truncation. In BP, a so-called neoclassical formation like *hipódromo* 'hippodrome' has its origin in Greek ἱππόδρομος 'circuit for chariot racing', but *-dromo* has the attribution of meaning 'to run', that is only diachronically possible (↗13 Morphology). In a modern synchrony, it was reinterpreted as 'place' (in fact, the hyperonym of the word) and henceforth it was possible to build other words like *sambódromo* 'sambadrome' (this BP word is now present in other languages), *camelódromo* 'official peddler area', *putódromo* 'official prostitution area',

protestódromo 'official area for protests'. The same process is valid for prefixes: from the compositional meaning of *transexual* 'transsexual' it was possible, by means of the truncation *trans* 'transsexual', to create new words like *transfobia* 'transphobia' (which is in fact a loan word from English) and *transfilia* 'transphilia' (idem). If a truncation is marked by the symbol |, we can freely say that a word like *foto* 'photo' is a truncation of *fotografia* 'photography', but etymologically speaking, its formula is: French PHOTO|GRAPHIE| → French PHOTO ▸ Portuguese *foto*. Every formula is an etymological *hypothesis*, which requires examination, and hence other solutions would be also plausible, such as: French PHOTOGRAPHIE ▸ Português FOTO|GRAFIA| → *foto* (Bizzocchi 1998; 2013). Only a careful examination of documents (and not individual introspection) would yield reliable results regarding the truth of each formula (Viaro [1]2011a). Neoclassical formations can be the source of new morphemes in derivation and composition: *foto-* could be associated both with 'light' (only by diachronic investigation, and is the origin, but not the etymon) and with 'photo' in words like *fotonovela* 'photo soap opera in magazines'. The same occurs with *heliporto* 'heliport', which is a loan word from English, since *heli-* is associated with *helicóptero* 'helicopter' and not with a propeller (from where *helico-* comes). In *fotonovela*, the morpheme *foto-* has a direct association with *fotografia*, and the word *foto* is indeed an independent lexical item, but in *heliporto*, even though a similar direct association with *helicóptero* exists, ∗*heli* is not an independent lexical item (nonexistent words must be marked by the symbol ∗ to avoid ambiguity in historical studies with the symbol *, which has been used since the 1850s got reconstructed forms, cf. Viaro [1]2011a; Viaro/Bizzocchi 2016). This latter process is very often used in splinters formations, like *paitrocínio* 'sponsorship by the father's money', *mãetrocínio* 'sponsorship by the mother', in which the splinter *-trocínio* evokes the word *patrocínio* 'sponsorship' (Gonçalves 2019, 155). A formula for a word like *paitrocínio* would be *pai* → *paitrocínio* ← |*pa*|*trocínio*.

Portmanteau words are clear examples of the need for the surrounding representation in etymological formulae within historical morphology studies. Theoretical problems can be more easily raised if portmanteau words are represented as follows: English SHOW ▸ *showmício* 'a rally with shows' ← Portuguese |CO|MÍCIO 'rally'. The same can be said of formulae of partial loanwords: Old French CHEVALIER ▸ Portuguese *cavaleiro* 'knight' ← Portuguese CAVALO 'horse'.

Merging etymological information is important when analogy is involved (a proposed symbol for analogy is ⇨, see Viaro [1]2011a; Viaro/Bizzocchi 2016). A classic example is Latin *FORESTA > Portuguese *floresta* 'forest' ⇦ FLOR 'flower'. The formula indicates, by means of an analogical hypothesis, that the diachronic phonetic change was not regular and needs explanation. It also indicates that *floresta* is not a morphological derivation from *flor*. Large-scale diachronic problems are accounted for by analogy, but the convincing power of formulae depends mainly on the reconstruction of entire past synchronies. We can say that *pedregoso* 'stony' comes from PEDRA 'stone' but the interfix *-eg-* is not easily explained. If we think of *-eg-* as a form of the Latin diminutive suffix *-ic-* that lingered through history, we can represent the formula as:

Portuguese PEDRA → Portuguese *pedreg*(*oso*) ⇦ Portuguese -EG- < Latin -IC-. The verisimilitude of this formula is dependent on the existence of other words with the *-eg-* suffix in the same past synchrony in which *pedregoso* was formed. Those words must either have been very frequent, to cause an analogical influence, or have belonged to a very large formal paradigm (which might now enjoy lower representation), in other words, to a prolific paradigm. Arguments will increase or decrease the level of certainty of a given formula. One can say that the frequent use of the combined ending *-egoso* increases the level of certainty. Therefore, influence is determined by frequency of use or paradigmatic prolificity in a given synchrony. Obviously, these conditions reveal a large number of problems to be resolved.

A formula such as Portuguese SABER ⇨ Portuguese *sabich*(*ão*) ⇦ Portuguese -ICH- is not interpreted as a solution for the formation problem of *sabichão* 'smart alec', but as a hypothesis whose level of certainty must be determined (Viaro/Bizzocchi 2016). That formula is better than Portuguese SABER → Portuguese *sabich*(*ão*) ⇦ Portuguese -ICH-, because the infinitive *saber* 'to know' is only an inflection with the ending morpheme *-er*. The infinitive, however, is the traditional form, adopted in lexicography, to evoke the entire verbal paradigm (*sabes*, *sabíamos*, *sabemos*, and so on), and thus must be understood as a label for the whole verbal paradigm (in other words, a hyperonym). But there are also other allomorphic stems for the same verb (*sei*, *saiba*, *soube*, *soubesse*) not directly used in the explanation of the derived form *sabichão*. That is why an analogical representation in such cases is preferable in order to avoid ambiguity.

Reminiscent analogies can complicate synchronic morphological analyses. Interfixes are a particular problem (Malkiel 1958). A word like *canavial* 'sugarcane plantation' has neither a *can-* base nor an *-avial* allomorph of *-al*. Because interfixes can be meaningless in a subsequent synchrony, they have always posed a serious problem for structuralist analysis if the solution of a synchronic segmentation is *can-avi-al*, because *-avi-* is a totally meaningless segment, with no direct or associative attribution to anything. An attempt at the historical analysis of this specific case would be an analogical explanation in a past synchrony, such as: Portuguese CANA → *canavial* ⇦ CÂNAVE 'hemp' < Latin CANNABEM.

4 Word classes in Brazilian Portuguese

To express the issue in another way, in historical morphology the problem of non-productive morphological elements such *heli-* in *heliporto* 'heliport', *-eiro* in *chuveiro* 'shower' and *-avi-* in *canavial* 'sugarcane plantation' is a problem of set theory. Cardinality is the number of elements in a set. Paradigms are nets of mental association, which can be represented as sets with a specific cardinality (called prolificity in morphology, as in Viaro 2010a). The phenomena mentioned in the previous section present the morphological base and means on which specific BP neologisms are built. Inherited from the same source, we could say that EP differs from BP only in terms of

frequency and cardinality. There are phonological, morphological, lexical and even syntactic paradigms. The label for these paradigms is the hyperonym. Some paradigms are singletons, with one element. In the morphology of BP, the paradigm *-avi-* is represented only by one lexical item, i.e., *canavial*. Diachronically speaking, some paradigms can still be prolific while no longer productive; others are prolific and also productive, with an indefinite cardinality. The resurrection of a no-longer-productive set is very common, and this means that productivity is rather unpredictable: the suffix *-eria*, as used to denote a business which prepares or serves food (attributed by the stem), became stagnant, only to be revigorated recently, as shown in BP words like *temakeria* 'a business which serves *temaki*', *boleria* 'a cake shop', *cupcakeria* 'a cupcake business'. So, a strict division between morphology and lexicology based on productivity is not always clear (certainly not as clear as supposed in Aronoff 1976).

Morphological information is not always as evident in a lexical item as other semantic information from the stem or from the affixes can lead us to imagine: a word like *menina* 'girl' is obviously feminine because it is possible to say that *-a* is a morpheme that indicates feminine morphological gender (and also referential sex or sexual identity) in opposition to *-o* of *menino* 'boy'. These morphemes can be segmented as *menin-a* and *menin-o*. But we also can say that *mesa* 'table' is a feminine word in Portuguese because of the ending *-a*, which has no referential sex, but the same morphological gender as *menina* could be represented as *mes:a*, thus being different from symbolic morphemes such as those like *menin-a*. In fact, the *-a* of *menina* is a symbolic sign, as shown in its opposition to the *-o* of *menino*, but the *-a* of *mesa* is an index (which could be notated as *mes:a* in order to distinguish it from the symbolic morpheme). Symbolic oppositions do not always involve sex or sexual identity, but there are lots of similar oppositions between referentially inanimate beings, as with *jarr-a*/*jarr-o* 'jar', *canec-a*/*canec-o* 'mug', *cest-a*/*cest-o* 'basket' whose semantical distinction exists, but it is not unanimously differentiated among Brazilian speakers, as a simple test illustrates: some people would say that a *jarra* is less sophisticated than a *jarro*, others will affirm that we cannot put flowers in a *jarra*, and yet others might claim that the material a jar is made from is crucial in determining the difference between a *jarra* and a *jarro*, and so on. We can affirm that the endings *-a* and *-o* form sets of words with a huge cardinality in the minds of Portuguese speakers. Let us label this set of 'feminine nouns ending in -a' and 'masculine nouns ending in -o', despite the existing feminine referentiality of these words. These big paradigms are also productive, and this is a different situation from the paradigm of the masculine nouns ending in *-a* (except *-ma* or *-ema*), which is unproductive. Masculine words ending in *-a*, such as *mapa* 'map', have neither a symbol nor an index, only a phoneme /a/.

In the word *parede* 'wall', no index is evident: the information of the feminine gender in this word is somewhat mixed in terms of meaning, because the nominal thematic vowel *-e* is not a symbolic morpheme or an index, but is also only a phoneme and an *-e* ending, used nowadays in BP as an optional expedient for inclusive (especially written) language: *alunes* 'male and female pupils/ students' (Lacotiz 2020). We

could say that morphological and semantic information are placed in the same portion of the linguistic sign, i.e., the signified (and not in the signifier or in the reference). If semantics is the study of the signified, morphology could be understood as a set of semantics *lato sensu* and distinct from those parts of the signified called semantics *stricto sensu*, or lexicology (Viaro 2018). In fact, morphological information is not a matter of the signifier, like phonemes or syllables, but has a special meaning, not a referential one like lexical meaning, which can be called morphological meaning. A form such as -AM is a typical ending for the singular accusative case of Latin first declension for nouns and adjectives, as in *puellam* 'girl' (accus.). It is also known that *puellam* is feminine and *nautam* 'sailor' is masculine; *puellam* is a noun and *pulchram* 'beautiful (fem. accus.)' is an adjective. Where is this information held? It is surely not in the signifier or in the reference. The only remaining choice is the signified (Coseriu [2]1987, 128–147). Thus, we can affirm that, in the realm of the signified, it is useful to separate lexical from morphological information. Other approaches, neglecting the existence of phenomena explainable only by a morphology *stricto sensu*, would seek to merge morphology with phonology or with syntax, in order to understand that morphology is a part of a morphophonology *lato sensu* or of a morphosyntax *lato sensu*. The signified, however, would deal with reference, whereas the signifier would not. Strictly morphological phenomena such as gender, number, case, declension, tense, mood, aspect, conjugation, person can indeed be paradigms of languages with no or little commitment to reference; morphological inflections are a kind of obsession of the *Weltanschauung* of the language (and not of the language's speakers or of the culture). They are subconscious glasses or 'straightjackets' without which social references of reality could not be discussed or even thought.

Symbolic morphemes and indexes can be changed diachronically in order to acquire redundant expression of morphological gender, or through expressiveness. Said Ali (1921; 1923) notes that today's masculine nouns *planeta* 'planet', *cometa* 'comet', *mapa* 'map', *fantasma* 'ghost' occur as feminine in old Portuguese, and the reverse also occurs with the current feminine noun *tribo* 'tribe': in those cases, the phoneme in the ending turned into a morpheme, acquiring morphological meaning, and has become an index. A more recent case is *grama* 'gram', originally masculine but nowadays frequently used as feminine in BP. Describing the expression of some Brazilian LGBT people, Vip/Libi (2006) cites interesting nouns whose indexes turn to symbolic morphemes in order to build words like *cabela* 'hair of a LGBT', *beiça* 'labia majora of the vagina', *crocodila* 'gossip girl', *mamífera* 'sexually insatiable woman', *ornitorrinca* 'woman, who hates sexually insatiable women', *prédia* 'building, where there is LGBT dating', *sapatã* 'novice lesbian' (their masculine stylistically non-marked counterparts are *cabelo* 'hair', *beiço* 'lip', *crocodilo* 'crocodile', *mamífero* 'mammal', *ornitorrinco* 'platypus', *prédio* 'building', *sapatão* 'lesbian'). A word that was motivated by reference is *personagem* 'character', originally only feminine (like other words in *-agem*, as mentioned above) and nowadays also masculine in BP like non-redundant forms such as *estudante* 'student'. In diminutive words it is possible to observe some

oscillation in words with no indexed ending like: (a) no redundance in: *problema* 'problem' (masc.) → *probleminha* 'little problem' (masc.); *pijama* 'pyjamas' (masc.) → *pijaminha* 'little pyjamas' (masc.); (b) emerging redundance in: *moto* 'motorcycle' (fem.) → *motinho* (fem.) ≈ *motinh:a* (fem.) 'a little motorcycle', but *peste* 'plague' (fem.) → *pestinh:a* 'little brat' (masc. or fem., but not ∗*pestinh:o*), *nenê* 'baby' (masc. or fem.) → *nenezinh:o* (masc.) and *nenezinh:a* (fem.).

Morphological number also sees oscillations in BP because of the eventual absence or presence of referential plurality, as in *costas* (pl.) ≈ *costa* (sing.) 'back', *óculos* (sing. or pl.) 'glasses', *diabete* (sing. masc or fem.) ≈ *diabetes* (pl. masc. or fem.) 'diabetes'. This is the same process found from ancient to modern Portuguese, as in *narizes* (pl.) → *nariz* (sing.) 'nose' in a past synchrony of the Portuguese language (Said Ali 1923).

We can say that an item like *mesa* 'table' belongs to lexical paradigms in which there are also items like *cadeira* 'chair', *sofá* 'sofa', *armário* 'cupboard', labelled with the hyperonym *móvel* 'furniture'. But the same words belong to the morphological paradigm whose label is "noun". This is a *morphological class*. That class has subsets like "feminine nouns" (*mesa*, *cadeira*) and "masculine nouns" (*sofá*, *armário*). A noun is different from an adjective because the items in its paradigm have morphological information in common (in fact, their label or hyperonym), which connects that noun hierarchically with other classes, say "subordinated to verbs" or "subordinator of adjectives"; an adjective also has the morphological information "subordinated to nouns". That information, which could be called *ranking index*, connects syntax and morphology. For instance, some items are present in both paradigms: *japonês* 'Japanese' can be a noun as well an adjective, but is a masculine noun (the corresponding feminine form would be *japonesa*).

Historically, it is possible to detect change in the ranking indexes leading to what traditionally used to be called 'improper derivation' (Port. *derivação imprópria*), also known by morphologists as 'conversion' (Port. *conversão*): the lexical items *sol* 'sun' and *trânsito* 'traffic' are nouns, but in BP it is common to find sentences like *tá trânsito na Avenida Paulista* 'there is traffic jam in Paulista Avenue' or *aqui tá sol* 'here it is sunny'. In fact, such adjectival use of *sol* in *tá sol* comes from the sentence *está fazendo sol* and reveals a speech change of the ranking index of the word *sol*, contextually understood as 'sunny' and not as 'sun' (Viaro 2018) and the relevant morphosyntactic information could be understood as present in the signified (as shown in the ubiquitous translation phenomenon by Tesnière 1959, 64).

Like the index range, another form of morphosyntactic evidence for the existence of gender information inside the signified of a morpheme is agreement, as usually shown in syntagms with a definite article: the word *estudante* belongs to both masculine and feminine noun paradigms, but sometimes that is only clear by means of agreement: *a estudante* 'the female student', *o estudante* 'the male student'. A noun can also be either masculine or feminine, but there is no change in the reference: in BP *alface* 'lettuce' is used in both gender agreements without meaning changes: *o alface*

and *a alface* (similar cases: *dó* 'pity', *cal* 'lime'). Morphological classes are therefore radical decisions of the signified with or without a connection to reference. A word like *criança* 'child' can be referentially a boy or a girl, yet lexical information is only 'a non-adult human being'; morphologically, it is always a feminine noun. The same can be said of *vítima* 'victim' and *testemunha* 'witness'. A word like *mulherão* 'a gorgeous woman' is a masculine noun in Portuguese, despite feminine reference. Another morphological class is verbal conjugation: lexical items like *cantar* 'to sing', *olhar* 'to look', *odiar* 'to hate' belong semantically to very different lexical paradigms, but to the same morphological one: they are infinitives of verbs of the so-called first conjugation. Classes are conceptual categories of the mind of the speaker and they should not be confused to the so-called 'word classes' of traditional grammar. Thinking so, irregular forms and exceptions are in fact only paradigms of low prolificity (i. e., sets with low cardinality). In historical morphology, a paradigm which is no longer productive can have high prolificity (like the verbs ending in *-ir*).

A suffixed word belongs to a particular morphological paradigm and its connections with lexical paradigms are not always evident. The Portuguese suffix *-eiro* is in fact a label (like the hyperonyms of the class names) of a very large set of nouns and adjectives. A description of the sign *-eiro* tells us that its signifier is [ˈejɾu] and its signified has a lot of lexical information such as "profession" and morphological information such as "masculine noun". The word *pedreiro* 'mason' is a derived noun from *pedra* 'stone', which is a feminine noun (and not a masculine one, like the derived noun). The hyperonyms of the lexical paradigms of *pedra* and *pedreiro* are also quite different (*pedra* is a mineral, and *pedreiro* is a profession) and some kind of implicit pragmatic information is needed to understand synchronically what a mason has to do with stones. If we say that *pedreiro* is a "man", this information has a referential descriptiveness that we call "meaning", which is evident by the ending *-o* (*pedreira* means not a female mason but a quarry). But the morphological information "masculine" is a signified without such a type of referential meaning in other words ending in *-eiro*, like *chiqueiro* "pigpen'. Its ranking index "noun" gives information about its agreement and subordination in the syntagmatic axis: lexical items ending in *-eiro* can be nouns, adjectives or both.

A Portuguese word of the form x-*eiro* can be interpreted by means of a paraphrase like '(h) which (v) x', in which h is the hyperonym (or the Aristotelean γένος), v is an implicit verb, deciphered only by means of pragmatics, and x is the *lexical base* or the *stem*. One can use that symbol exclusively for nominal stems and x^V for verbal ones (Rio-Torto 1998; Viaro 2013a). In paraphrases it is necessary to avoid vague or ambiguous hyperonyms such as "thing", "act", "product", in order to establish diachronic meaning dependences. The hyperonym h can be variable and eventually indefinite, e.g., 'person who x^V', 'object that x^V', 'which lives in x', 'which comes from x', 'which follows x's teachings', 'which belongs to x', 'bad x', 'a great quantity of x', 'which resembles x', 'typical of x', 'which x^v a lot', 'which often visits x', 'person who deals with x', 'which x^V', 'person who thinks like x' and that depends on lexical signified of x. For

instance, if x is associated with naming a fruit (*abacat-* 'avocado'), the hyperonym of the derived lexical item is normally 'fruit tree,' and the interpretation of the paraphrase would be 'fruit tree which (produces) x': *abacat-eiro* 'avocado tree'. The absence of suffix analogies for the latter interpretation in BP characterizes the subset of x-*eiro* fruit tree names. Another peculiarity is the same gender used for both the noun associated with the base x and for the derived noun x-*eiro*: GOIABA (fem.) → *goiabeira* (fem.) 'guava tree', LARANJA (fem.) → *laranjeira* (fem.) 'orange tree', LIMÃO (masc.) → *limoeiro* (masc.) 'lemon tree'. All those conditions and information are contained inside the signified of the suffix and, in fact, the new suffixed words in *-eiro* for fruit plants, created in BP with respect to those conditions: *açaí* (masc.) → *açaizeiro* (masc.) 'açaí palm'; *mangaba* (fem.) → *mangabeira* (fem.) 'mangaba bush'.

The acceptance of the existence of undefined word classes (i.e., with no information about the ranking index of a language) has consequences in the taxonomy of the resultant words of a derivation and is a problem for historical morphology. Some suffixes are frequently described as "denominal", "deadjectival", "deverbal", "departicipial" or "deadverbial", but the base *pedr-* in *pedreiro* is precisely not a noun like PEDRA (although we can say that it was a noun from the diachronic point of view). In spite of this, *-eiro* in *pedreiro* is called a denominal suffix. The classification of other cases is even more difficult: *alfinetada* 'a jab (at)' is a feminine noun but its base *alfinet-* can be interpreted as a noun (ALFINETE 'a pin') or as a verb (ALFINETAR 'to pin').

5 Meaning from the meaningless

As discussed above with examples from BP, it is not uncommon that phonemes become morphemes, like indexes, and that indexes may become wholly functional and distinctive symbolic signs. If we represented diachronic meaning changes of a certain morphological component by means of a tree diagram (somewhat similar to the phylogenies in biology, as in Viaro 2011c, 150, 176), say, a derivational suffix, it would be possible to track the original changes in meaning step by step, frequently from a concrete to a more abstract and pejorative meaning (although the inverse is also common, as we see in the semantic transformation of the previously discussed Latin adjective x-ARIUS 'related to x' into more concrete tree denominating markers in x-*eiro* 'fruit tree which produces x'). We postulate an original affixed word with a global meaning formed by the sum of the meanings of the base and its affixes. From the first global lexical meaning, a semantic path can be created in subsequent synchronies. There are, however, examples of specific meanings which orbit every prototypical secondary meaning with which they are genealogically linked. Less productive variations such as 'tree which produces x' (*palmeira* 'palm tree') or 'plant which produces x' (*roseira* 'rosebush') are non-prototypical meanings of x as 'fruit', but historically residual lexical paradigms. The word *trepadeira* 'bindweed' has the same hyperonym ('plant'), but the first meaning of its suffix was completely different: 'which (often)

v' (the same can be said about *dormideira* 'sleepy plant', the stem of which is also departicipial).

The arbitrary relation between signifier and signified presented in Saussure (1916) and that of reference abandoned by most structuralists must inevitably be recalled in the discussion about assumptions within morphological theory. We might think that there were other special kinds of meaning related to understanding of parole, and not only to expression. Pejorativity, for example, can emerge from a syntactic interpretation and become a nuclear meaning of a suffix studied in historical morphology. The same occurs with poetic meaning, slang, and some idiosyncratic and contextual meanings. Those inter signa meanings can become expressionless clichés and migrate to the signified of the sign of the langue by means of historical change. It is common to talk about new meanings developed by metaphor or metonymy, but that movement is not very different from the formation ex nihilo of a new meaningful portion from a previous non-semantic sequence of phonemes (Viaro/Ferreira/Guimarães Filho 2013). That is common in historical morphology and to some degree defies dogmatic positions about the thesis of the arbitrary nature of the sign, as in the takete/baluba test in psychology (Köhler 1929). For example, the new Portuguese suffix -Vngo (i.e. *-ango, -engo, -ingo, -ongo, -ungo*), particularly productive in BP, merged derived meanings of an old suffix *-engo* (of Germanic origin) and an unclear pejorative potential for expression induced from neology and from real loanwords of non-Indo-European languages (especially Tupi and Bantu languages). Typically, those forms slowly migrated to familiar contexts from the colloquial language and have acquired frequently pejorative interpretations: Portuguese VERDE → Portuguese *verdolengo* 'green (and ugly)', Portuguese MULHER → Portuguese *mulherengo* 'womanizer', English CAR → BP *caranga* 'a cool car', Portuguese JAPO|NÊS → BP *japorongo* 'a Japanese (derogatory)', English HIPPIE → BP *riponga* 'a hippie (derogatory)'. It is also possible to talk about pejorative evaluation in a given synchrony detached from phonic patterns like –VnCo (if V is a vowel and C is an unvoiced plosive consonant). The gitaigo class in the traditional grammar of Japanese also exhibits problematic examples of association of meaning with such premorphemic unities of signifiers (Viaro 2005; Viaro/Ferreira/Guimarães Filho 2013).

6 Conclusions

If we understand a language as a political agreement among speakers who understand a certain variation in expression as belonging to the same object of knowledge, it is not heretical to say that a language is not a system, but a set of slightly (or even considerably) different systems. But if even tiny individual differences can be considered different systems, some kind of miracle happens when two individuals understand each other. Tolerance to that variability or a higher abstraction of those slight variants is not enough to say that the word "language" is a strictly linguistic term without political implications. The existence of a common origin prior to the 11th cen-

tury is not an argument in support of the assertion that to say that Galician and Portuguese are the same language, because the old Galician language was modified after conquering southern lands of future Portugal: Galicia did not dominate Portugal after the independence of Portugal, and the inverse didn't occur after that either. If we think on these lines, could we say that a peculiar BP morphological system exists after 1822? And, if this were the case, does the relative isolation conferred on Brazil by the Atlantic Ocean promote another type of reasoning, one that leads us to be able to safely shift the beginnings of that variant to the 16th century?

A large number of morphological structures present in BP were inherited from pre-16th century variants of EP. After that, we can say that two standard variants of the same language were gradually formed, beside many substandard variants, on both sides of Atlantic Ocean, and that these were culturally in contact until the 19th century, and indeed after that. What is typical of BP, and was not inherited from Portugal nor is used anymore in Portugal, is difficult to determine nowadays, and not only in morphology. There was a relatively uninterrupted relation between the two countries: EP words were adopted in Brazil, mainly when Brazil was a Portuguese colony; but the appearance of BP words in Portugal has also not been completely unknown in Portugal, this due to many factors (typical denominations of animals and plants, mutual economic relations and more recently soap operas and films) contributing to BP loan words in EP, many of them affixed words which contribute to the prolificity and productivity of the morphemes involved. Immigration in both directions also served to keep lexical and morphological transit alive. Obviously, there are differences, but the exhaustive list is unknown, because standard written language tends to minimize the real distance in speech of informal forms of expression. We have good diachronic descriptions of the Portuguese language, but reconstructions of past synchronies still need to be made. Past synchronies, however, are scant and often incomplete, not only in BP but indeed in most known languages, which makes that task even more difficult. This *sine qua non* demands urgency in the future schedule of diachronic studies.

7 References

Amaral, Amsadeu (1920), *O dialeto caipira*, São Paulo, O Livro.

Areán-García, Nilsa (2007), *Estudo comparativo de aspectos semânticos do sufixo "-ista" no português e no galego,* 2 vol., Master's thesis, São Paulo, FFLCH/USP, https://www.teses.usp.br/teses/disponiveis/8/8142/tde-29082012-103453/pt-br.php, last accessed 16.02.2022.

Areán-García, Nilsa (2011), *Aspectos sincrônicos e diacrônicos do sufixo "-ístico(a)" no português e no galego*, 2 vol., Doctoral thesis, São Paulo, FFLCH/USP, http://www.teses.usp.br/teses/disponiveis/8/8142/tde-04102007-144618/pt-br.php, last accessed 16.02.2022.

Areán-García, Nilsa (2013), *Algumas características do sufixo "-ista"*, in: Mário Eduardo Viaro (ed.), *Morfologia Histórica*, São Paulo, Cortez, 190–219.

Aronoff, Mark (1976), *Word formation in generative grammar*, Cambridge, Massachusetts/London, MIT Press.

Becker, Martin (2013), *O sufixo "-ada" em português*, in: Mário Eduardo Viaro (ed.), *Morfologia Histórica*, São Paulo, Cortez, 118–153.
Bizzocchi, Aldo Luiz (1998), *Léxico e ideologia na Europa Ocidental*, São Paulo, Anneblume/FAPESP.
Bizzocchi, Aldo Luiz (2013), *Processos de formação lexical das línguas românicas e germânicas: uma nova perspectiva teórica*, Domínios da Lingu@gem 7/1, 9–39, http://www.seer.ufu.br/index.php/dominiosdelinguagem/article/view/19495/12552, last accessed 16.02.2022.
Campos, Lucas S. (2004), *A negação prefixal na história da língua portuguesa*, Doctoral thesis, Salvador, PROHPOR/UFBA, https://repositorio.ufba.br/handle/ri/28698, last accessed 16.02.2022.
Coelho, Juliana Soledade (2000), *Aspectos morfolexicais do português arcaico (séculos XIII e XIV)*, Master's thesis, Salvador, PROHPOR/UFBA.
Coelho, Juliana Soledade (2005), *Semântica morfolexical: contribuições para a descrição do paradigma sufixal do português* arcaico, Doctoral thesis, Salvador, PROHPOR/UFBA.
Coelho, Juliana Soledade/Lopes, Mailson (2015), *Uma proposta de revisão do conceito de morfema*, in: Juliana Soledade Coelho et al. (edd.), *Para a construção de saberes lexicais*, Salvador, EDUFBA, 127–159, https://43f732ad-b5b4-4926-ad06-b21be791f1d4.filesusr.com/ugd/c8e334_b5879e3c9bc54a83a742ee32e1e8b5f2.pdf, last accessed 16.02.2022.
Cornu, Jules (1888), *Grammatik der portugiesischen Sprache*, in: Gustav Gröber (ed.), *Grundriss der romanischen Philologie*, Straßburg, Trübner, vol.1, 715–803.
Coseriu, Eugenio ([2]1987), *Gramática, semántica, universales: estudios de lingüística funcional*, Madrid, Gredos.
Cunha, Antônio Geraldo da (1982), *Dicionário etimológico Nova Fronteira da língua portuguesa,* Rio de Janeiro, Nova Fronteira.
Cunha, Antônio Geraldo da (2006), *Vocabulário histórico-cronológico do português medieval*, Rio de Janeiro, Casa de Rui Barbosa (CD-Rom).
Freitas, Érica Santos Soares de (2008), *Em busca do mento perdido. Análise semântico-diacrônica do sufixo "-mento" no português*, 2 vol., Master's thesis, São Paulo, FFLCH/USP, http://www.teses.usp.br/teses/disponiveis/8/8142/tde-29092008-151130/pt-br.php, last accessed 16.02.2022.
Freitas, Érica Santos Soares de (2013), *O sufixo "-mento" sob uma perspectiva diacrônica*, in: Mário Eduardo Viaro (ed.), *Morfologia Histórica*, São Paulo, Cortez, 243–262.
Freitas, Érica Santos Soares de (2014), *O tempo e o "-mento": história do sufixo latino "-mentum" e de seu desenvolvimento na língua portuguesa, em contraste com outras línguas românicas*, Doctoral thesis, São Paulo, FFLCH/USP, http://www.teses.usp.br/teses/disponiveis/8/8142/tde-22042014-125405/pt-br.php, last accessed 16.02.2022.
Freitas, Érica Santos Soares de (2018), *Glossário de palavras portuguesas sufixadas em "-mento"*, Doctoral thesis, São Paulo, FFLCH/USP, http://www.livrosabertos.sibi.usp.br/portaldelivrosUSP/catalog/book/220, last accessed 16.02.2022.
Gerards, David Paul/Kabatek, Johannes (2018), *Grammaticalization, distance, immediacy and discourse traditions: the case of Portuguese "caso"*, in: Salvador Pons Bordería/Óscar Loureda Lamas (edd.), *Beyond grammaticalization and discourse markers: new issues in the study of language change*, Leiden/Boston, Brill, 115–159.
Gianastacio, Vanderlei (2009), *A presença do sufixo "-ismo" nas gramáticas da língua portuguesa e sua abrangência dos valores semânticos, a partir do Dicionário de Língua Portuguesa Antônio Houaiss,* Master's thesis, São Paulo, FFLCH/USP, http://www.teses.usp.br/teses/disponiveis/8/8142/tde-30112009-151358/pt-br.php, last accessed 16.02.2022.
Gianastacio, Vanderlei (2013), *A origem do sufixo "-ismo": da língua grega à latina e desta à portuguesa*, in: Mário Eduardo Viaro (ed.), *Morfologia Histórica*, São Paulo, Cortez, 220–242.

Gianastacio, Vanderlei (2015), *A presença dos sufixos "-sc-" e "-iz-" na Vulgata e sua abrangência semântica e aspectual*, Doctoral thesis, São Paulo, FFLCH/USP, http://www.teses.usp.br/teses/disponiveis/8/8142/tde-29042015-175910/pt-br.php, last accessed 16.02.2022.

Gonçalves, Anielle Aparecida Gomes (2009), *Diacronia e produtividade dos sufixos "-agem", "-igem", "-ugem", "-ádego", "-ádigo" e "-ádiga" em português*, Master's thesis, São Paulo, FFLCH/USP, http://www.teses.usp.br/teses/disponiveis/8/8142/tde-30112009-142459/pt-br.php, last accessed 16.02.2022.

Gonçalves, Anielle Aparecida Gomes (2013), *Origem e análise semântica dos sufixos "-agem", "-igem", "-ádego", "-ádigo" e "-ádiga*," in: Mário Eduardo Viaro (ed.), *Morfologia Histórica*, São Paulo, Cortez, 263–282.

Gonçalves, Carlos Alexandre (2019), *Morfologia*, São Paulo, Parábola.

Heckler, Evaldo/Back, Sebald/Massing, Egon (1984), *Dicionário morfológico da língua portuguesa,* 5 vol. São Leopoldo, Unisinos.

Heckler, Evaldo/Back, Sebald/Massing, Egon (1994), *Estrutura das palavras*, São Leopoldo, Unisinos.

Houaiss, Antônio/Villar, Mauro (2001), *Dicionário Houaiss da língua portuguesa*, Rio de Janeiro, Objetiva.

Huber, Joseph (1933), *Altportugiesisches Elementarbuch*, Heidelberg, Winter.

Ilari, Rodolfo (ed.) (2014), *Palavras de classe aberta*, São Paulo, Contexto.

Ilari, Rodolfo (ed.) (2015), *Palavras de classe fechada*, São Paulo, Contexto.

Köhler, Wolfgang (1929), *Gestalt Psychology*, New York, Leveright.

Lacotiz, Andréa (2007), *Estudo diacrônico da função e dos valores semânticos dos sufixos "-ança"/"-ença"/"-ância"/"-ência" no português*, Master's thesis, São Paulo, FFLCH/USP, http://www.teses.usp.br/teses/disponiveis/8/8142/tde-28012008-112539/pt-br.php, last accessed 16.02.2022.

Lacotiz, Andréa (2013), *O surgimento de "-ntia" e sua disseminação nas línguas neolatinas*, in: Mário Eduardo Viaro (ed.), *Morfologia Histórica*, São Paulo, Cortez, 308–314.

Lacotiz, Andréa (2020), *Flexão de gênero: estudo historiográfico sobre a genealogia dos conceitos e abordagem semiótica da morfologia no português*, Doctoral thesis, São Paulo, FFLCH-USP, https://www.teses.usp.br/teses/disponiveis/8/8142/tde-06042021-145515/pt-br.php, last accessed 16.02.2022.

Leoni, Francisco (1858), *Genio da lingua portugueza ou causas racionaes e philologicas de todas as formas e derivações da mesma lingua, comprovadas com innumeraveis exemplos extrahidos dos auctores latinos e vulgares*, 2 vol., Lisboa, Typographia do Panorama.

Longhin, Sanderléia Roberta (2003), *A gramaticalização da perífrase conjuncional "só que",* Doctoral thesis, Campinas, Unicamp, https://core.ac.uk/download/pdf/296832901.pdf, last accessed: 16.02.2022.

Lopes, Celia Regina dos Santos (1998), *"Nós" e "a gente" no português falado culto do Brasil*, D.E.L.T.A. 14/2, 405–422.

Lopes, Celia Regina dos Santos (2003), *A inserção de "a gente"no quadro pronominal do português,* Frankfurt am Main/Madrid, Vervuert/Iberoamericana.

Lopes, Mailson dos Santos (2013), *A prefixação na primeira fase do português arcaico: descrição e estudo semântico-morfolexical-etimológico do paradigma prefixal da língua portuguesa nos séculos XII, XIII e XIV*, Master's thesis, Salvador, PROHPOR/UFBA, https://repositorio.ufba.br/ri/handle/ri/29879, last accessed 16.02.2022.

Lopes, Mailson dos Santos (2018), *Estudo histórico-comparativo da prefixação no galego-português e no castelhano arcaicos (séculos XIII-XVI): aspectos morfolexicais, semânticos e etimológicos,* Doctoral thesis, Salvador, PROHPOR/UFBA, https://repositorio.ufba.br/ri/handle/ri/15537, last accessed 16.02.2022.

Lucchesi, Dante (2009), *A concordância de gênero*, in: Dante Lucchesi/Alan Baxter/Ilsa Ribeiro (edd.), *O português afro-brasileiro*, Salvador, EDUFBA, 295–318.
Malkiel, Yakov (1958), *Los interfijos hispánicos*, in: Diego Catalán (ed.), *Estructuralismo e historia II: Miscelánea homenaje a André Martinet*, Tenerife, Universidad de la Laguna, 107–187.
Nunes, José Joaquim ([3]1943, [2]1921, [1]1906), *Chrestomathia archaica: excerptos da litteratura portuguesa desde o que mais antigo se conhece até ao seculo XVI acompanhados de introdução gramatical, notas e glossario*, Lisboa, Ferreira & Oliveira.
Nunes, José J. ([3]1945, [1]1919), *Compêndio de gramática histórica portuguesa*, Lisboa, Clássica.
Oliveira, Anielle Aparecida Gomes Gonçalves Jacometti de (2014), *Os sufixos agentivos "-nte" e "(d/t/s)or" no português: um estudo semântico-diacrônico*, Doctoral thesis, São Paulo, FFLCH/USP, http://www.teses.usp.br/teses/disponiveis/8/8142/tde-24032015-120314/pt-br.php, last accessed 16.02.2022.
Pharies, David (2002), *Diccionario etimológico de los sufijos españoles y de otros elementos finales*, Madrid, Gredos.
Rio-Torto, Graça Maria (1993), *Formação de palavras em português: aspectos da construção de avaliativos*, Doctoral thesis, Coimbra, Faculdade de Letras da Universidade de Coimbra.
Rio-Torto, Graça Maria (1998), *Morfologia derivacional: teoria e aplicação ao português*, Porto, Porto Editora.
Rio-Torto, Graça Maria (2013), *Desafios em morfologia: história e (re)conhecimento*, in: Mário Eduardo Viaro (ed.), *Morfologia Histórica*, São Paulo, Cortez, 31–57.
Rocha, Luís Carlos de Assis (1999), *A nominalização no português do Brasil*, Revista de Estudos Linguísticos 8/1, 5–51, http://www.periodicos.letras.ufmg.br/index.php/relin/article/view/2301/2250, last accessed 16.02.2022.
Rodrigues, Angela/Alves, Ieda Maria (edd.) (2015), *A construção morfológica da palavra*, São Paulo, Contexto.
Said Ali, Manoel (1921), *Lexeologia do portuguez historico*, São Paulo, Melhoramentos.
Said Ali, Manoel (1923), *Formação de palavras e syntaxe do portuguez histórico*, São Paulo, Melhoramentos.
Santana, David de Oliveira (2008), *Prefixos derivados de preposições em textos do século XVII ao XX*, Master's thesis, Salvador, PROHPOR/UFBA, https://repositorio.ufba.br/ri/handle/ri/11640, last accessed 16.02.2022.
Santana, Messias dos Santos (2017), *O sufixo diminutivo em português. Forma, funcionamento e significação*, Doctoral thesis, São Paulo, FFLCH/USP, http://www.teses.usp.br/teses/disponiveis/8/8142/tde-13062017-161624/pt-br.php, last accessed 16.02.2022.
Santos, Alice Pereira (2010), *Polissemia dos sufixos aumentativos "-ão", "-arro", "-orro", "-aço" e "-uço" e seus traços avaliativos sob a perspectiva diacrônica*, Master's thesis, São Paulo, FFLCH/USP, http://www.teses.usp.br/teses/disponiveis/8/8142/tde-11102011-101958/pt-br.php, last accessed 16.02.2022.
Santos, Alice Pereira (2013), *Estudo semântico-histórico do sufixo "-udo"*, in: Mário Eduardo Viaro (ed.), *Morfologia Histórica*, São Paulo, Cortez, 283–307.
Santos, Alice Pereira (2016), *Morfologia em diacronia—os caminhos e desvios de um afixo na história da língua: o percurso histórico-semântico do prefixo "des-" em bases sufixadas e em formações parassintéticas*, Doctoral thesis, São Paulo, FFLCH/USP, https://www.teses.usp.br/teses/disponiveis/8/8142/tde-10052016-115122/pt-br.php, last accessed 16.02.2022.
Santos, Antônia Vieira dos (2009), *Compostos sintagmáticos nominais VN, NN, NA, AN e NprepN no português arcaico (séculos XIII a XVI),* Master's thesis, Salvador, PROHPOR/UFBA, https://repositorio.ufba.br/ri/handle/ri/11221, last accessed 16.02.2022.
Saussure, Ferdinand de (1916), *Cours de linguistique générale*, Lausanne/Paris, Payot.

Simões Neto, Natival Almeida (2016), *Um enfoque construcional sobre as formações "X-eir-": da origem latina ao português arcaico*, Master's thesis, Salvador, PROHPOR/UFBA, https://repositorio.ufba.br/ri/handle/ri//30404, last accessed 16.02.2022.

Simões Neto, Natival Almeida (2019), *O esquema "X-ari" do latim às línguas românicas: um estudo comparativo, cognitivo e construcional*, Doctoral thesis, Salvador, PROHPOR/UFBA, https://repositorio.ufba.br/ri/handle/ri/32014, last accessed 16.02.2022.

Takahashi, Mônica Yuriko (2013), *Mudanças semânticas do sufixo "-ada"*, in: Mário Eduardo Viaro (ed.), *Morfologia Histórica*, São Paulo, Cortez, 335–350.

Tesnière, Lucien (1959), *Éléments de syntaxe structurale*, Paris, Klincksieck.

Viaro, Mário Eduardo (2005), *Relação entre produtividade e freqüência na produção do significado*, Estudos Lingüísticos 34, 1230–1235, http://www.usp.br/gmhp/publ/Via29.pdf, last accessed 16.02.2022.

Viaro, Mário Eduardo (2007), *Estudo diacrônico da formação e da mudança semântica dos sufixos "-eiro"/"-eira" na língua portuguesa*, in: Gladis Massini-Cagliari et al. (edd.), *Trilhas de Mattoso Câmara e outras trilhas: fonologia, morfologia, sintaxe*, São Paulo, Cultura Acadêmica, 45–84, http://www.usp.br/gmhp/publ/via33.pdf, last accessed 16.02.2022.

Viaro, Mário Eduardo (2010a), *Sobre a inclusão do elemento diacrônico na teoria morfológica: uma abordagem epistemológica*, Estudos de Lingüística Galega 2, 173–190. http://www.usp.br/gmhp/publ/ViaA4.pdf, last accessed 16.02.2022.

Viaro, Mário Eduardo (2010b), *A especialização do sufixo latino "-arium"*, in: Maria João Marçalo et al. (edd.), *Língua portuguesa: ultrapassar fronteiras, juntar culturas*, Évora, Universidade de Évora, 22–42, http://www.usp.br/gmhp/publ/ViaA2.pdf, last accessed 16.02.2022.

Viaro, Mário Eduardo ([2]2014, [1]2011a), *Etimologia*, São Paulo, Contexto.

Viaro, Mário Eduardo (2011b), *A formação do significado agentivo de "-eiro"*, in: *Actas del XVI Congreso Internacional de la Asociación de Lingüística y Filología,* Alcalá de Henares, Universidad de Alcalá, 2671–2679, http://www.usp.br/gmhp/publ/ViaA5.pdf, last accessed 16.02.2022.

Viaro, Mário Eduardo (2011c), *A derivação sufixal do português: elementos para uma investigação semântico-histórica*, Habilitation thesis, São Paulo, FFLCH/USP, http://www.teses.usp.br/teses/disponiveis/livredocencia/8/tde-02022018-173614/pt-br.php, last accessed 16.02.2022.

Viaro, Mário Eduardo (2012), *Linguística da comunicação e linguística descritiva: os eixos sincrônico e diacrônico nos atuais modelos da morfologia*, Estudos Linguísticos 41, 277–290, http://www.usp.br/gmhp/publ/ViaA7.pdf, last accessed 16.02.2022.

Viaro, Mário Eduardo (ed.) (2013a), *Morfologia Histórica*, São Paulo, Cortez.

Viaro, Mário Eduardo (2013b), *A morfologia histórica e os estudos etimológicos da língua portuguesa*, Revista Diadorim, nº especial, 39–64, DOI: https://doi.org/10.35520/diadorim.2013.v0n0a4006.

Viaro, Mário Eduardo (2018), *Onde está a morfologia no signo linguístico?*, Estudos linguísticos e literários 61, 11–29, https://periodicos.ufba.br/index.php/estudos/article/viewFile/28109/19115, last accessed 26.02.2022.

Viaro, Mário Eduardo/Bizzocchi, Aldo Luiz (2016), *Proposal of new concepts and a new notation in formulating etymological propositions and discussions*, Alfa 60/3, 579–601, http://www.scielo.br/pdf/alfa/v60n3/en_1981-5794-alfa-60-3-0579.pdf, last accessed 16.02.2022.

Viaro, Mário Eduardo/Ferreira, Michael/Guimarães Filho, Zwinglio (2013), *Derivação ou terminação: limites para a semântica, lexicologia e morfologia históricas*, in: Mário Eduardo Viaro (ed.), *Morfologia Histórica*, São Paulo, Cortez, 58–105.

Vip, Ângelo/Libi, Fred (2006), *Aurélia: a dicionária da língua afiada*, São Paulo, Editora da Bispa.

Williams, Edwin (1938), *From Latin to Portuguese: historical phonology and morphology of the Portuguese language*, Oxford, Oxford University Press.

Charlotte Galves and Célia Regina dos Santos Lopes

6 Historical syntax

Abstract: This chapter presents the main morpho-syntactic changes that have occurred in the history of Brazilian Portuguese which led to the emergence of a new syntax, different from both Modern European Portuguese and the ancestral language, Classical Portuguese. Relying mainly on written documents, but also on more recent oral corpora, we first discuss the reorganization of the pronominal paradigm and its consequences for both the morpho-syntax of clitics and the occurrence of null arguments. Then, we analyze the morphological and syntactic changes in verbal constructions. Finally, we study syntactic phenomena involving subjects and topics as well as the peculiarities of WH-sentences, which distinguish Brazilian from European Portuguese. Considerations about the dynamics of these changes are provided over the course of the paper.

Keywords: Brazilian Portuguese, Classical Portuguese, syntactic change, address forms, clitic-placement, null subjects, periphrastic future, existential constructions, subject-verb order, topic-subjects, WH- clauses

1 Introduction

This chapter focuses on the diachronic changes that led to the emergence of a new syntax in the variety of Portuguese spoken in Brazil. In the last twenty years, a large amount of studies have been performed in order to produce documents that can help researchers better understand and describe the nature and dynamics of such changes. Thanks to the work of large collective projects, in which thousands of documents from all around the country have been collected, edited, and annotated, leading to the production of corpora,[1] books, papers, and dissertations, the picture has gradually gained more precise contours. As a result, previous descriptions based on sparser or less systematic data can be enriched and revised. Here, we give an overview of the main results of this considerable work, aiming to highlight, when possible, the complex and intricate paths of the evolution of the main syntactic phenomena that distinguish Brazilian Portuguese (BP) from Modern European Portuguese (EP). We say "when possible" because not all phenomena have been documented in the historical

1 Cf. the History of Brazilian Portuguese (PHPB) Project and the Tycho Brahe Parsed Corpus of Historical Portuguese Project: http://www.tycho.iel.unicamp.br/~tycho/corpus/en/index.html, last accessed 16.02.2022.

https://doi.org/10.1515/9783110405958-007

records. This may be due to the normative constraints operating on written texts, even when they are written in familiar contexts.

Additionally, certain phenomena that we know are peculiar to contemporary BP may have emerged in speech later than others and have only begun to appear in texts much more recently. For a complete overview of the syntactic properties of BP, we refer the reader to ↗14 Syntax. Two important issues have been raised in connection with the diachronic studies of BP syntax. One regards the nature of the language that was brought to Brazil by the first colonizers (Ribeiro 1998; Galves in prep.). Studies on the history of EP show that the language written, and possibly spoken, in the 16th and the 17th centuries—the period in which Brazil was colonized by the Portuguese—is different from both Old Portuguese and EP. Following tradition, we shall call it "Classical Portuguese". We assume that this is the language that is at the roots of BP.[2] From this point of view, when Modern European Portuguese (what we have been calling EP), which emerged in Portugal in the first half of the 18th century,[3] arrives in Brazil, at a time of intense emigration due to the gold rush, the Brazilian variety is already sufficiently strong to not suffer a deep influence from it. However, when the question of a Brazilian standard emerges in the second half of the 19th century, the influence of EP is visible in texts. Additionally, the constant arrival of speakers of EP throughout the entire colonial period played a major role in preventing the morphology of BP from moving too far away from its Portuguese roots (cf. Lucchesi/Baxter/Ribeiro 2009).

The second important question is whether the morphosyntactic changes that occurred in Brazil are due to the natural evolution of languages over time, or whether they result from the massive contact process due to the social history of Portuguese in America. This dichotomy has long occupied the linguistic field in Brazil and is still found in recent works such as Naro/Scherre (2007). The sections that follow do not address this question directly. However, we believe that for the historical reasons presented in ↗2 The social history of Brazilian Portuguese and ↗3 The history of linguistic contact underlying Brazilian Portuguese, the extensive linguistic contact in Brazil, mainly during the colonial period, must have played a decisive role in the emergence of a different syntax along the lines described in this chapter. This does not mean, as has been frequently assumed, that BP underwent creolization followed by a subsequent decreolization (cf. Guy 1981 and the discussion in Tarallo 1993), which has yet to be supported by any historical record. What is instead implied is that the changes described in the following sections are the result of a complex interplay of antagonistic forces, which can be summarized in linguistic terms by the interaction of the transmission of Portuguese both as a first and second language over a long period of time (cf. Mattos e Silva 2004; Lucchesi/Baxter/Ribeiro 2009; Avelar/Galves 2014; and ↗3 The history of linguistic contact underlying Brazilian Portuguese).

2 For a different point of view, see Castilho (2013).

3 cf. Paixão de Sousa (2004); Galves/Britto/Paixão de Sousa (2005); Galves/Paixão de Sousa (2017).

This chapter is organized as follows. In Section 2, we describe the reorganization of the pronominal paradigm and its consequences on the syntax of clitics and the occurrence of null arguments. At the end of the section, we discuss the abstract nature of such changes. Section 3 is devoted to the morphological and syntactic changes undergone by verbal constructions. Finally, Section 4 addresses issues concerning the evolution of the syntax of clauses, from both the inside—focusing on phenomena of order, case, and agreement—and the outside—looking at special types of clauses that have had distinct relevance in the syntactic history of BP. Section 5 briefly closes the chapter.

2 Personal pronouns: morpho-syntax and use

2.1 A new pronominal system: changes in forms of address

The set of personal pronouns in BP presents innovations that do not appear in EP. The differences occur not only in subject position (nominative) but also for direct (accusative) and indirect (dative) complements as well as the possessive (genitive). The gradual emergence of new forms from nominal expressions led to a series of effects in different aspects of Portuguese grammar particularly in Brazil. This happened in the first person plural with the rise of *a gente* 'we'— originating from the noun *gente* 'people' (cf. Rumeu 2013; Lopes 2003; and others) and in the second person (singular and plural), as we discuss here.

The origin of the changes in the pronominal paradigm of the 2nd person is related to the changes, over time, in the forms of address to express politeness and intimacy. Until the 15th century, the opposition between intimacy and politeness occurred between *tu* 'you, SG' and *vós* 'you, PL'—the latter, originally 2PL, extended to the 2SG. From the 15th century on, the address form *Vossa Mercê* 'Your Grace,' among others, was used together with *vós* ('you.PL.NOM' for 2SG.NOM) to show reverence to the king (cf. Cintra 1972). The semantic deterioration of address forms followed social changes in Portuguese society: *Vossa Mercê* 'Your Grace' or 'Your Mercy' lost its early polite use when it was gradually extended to other social layers, such as the nobles at first and the rising bourgeoisie later, which began to demand a respectful address.

From the 16th century on, the address form *Vossa Mercê* and its variants *vosmecê*, *mecêa, vosse*, and *você* 'you.SG' began to be used in the same functional domain as variant forms in asymmetrical relations (cf. Lopes/Duarte 2003, 65).

From the mid-18th century onwards, *Vossa Mercê* and *você* became functionally and discursively divergent. The former continued to be used in ascending asymmetrical relations (1), while the new form was frequent in descending relations (2):

(1) ***Vossa Mercê*** ha de ter susego só quando receber esta.
'**Your Grace** (2SG.NOM) will calm down only when you receive this (letter)'
(Letter written from son to mother, 19th century)

(2) Minha Filha. Todos os dias espero receber carta sua [...] ***voce*** a muito não escreve.
'My daughter. Everyday I hope to receive a letter from you [...] **you** (2SG.NOM.V) haven't written for a long time'
(Letter written from mother to daughter, 19th century, in Fraga/Gomes 2014)

Until this period, the pronominal system of EP and BP was equivalent. From the 19th century on, unlike what was happening in Portugal, the new form *você* 'you.SG' began to be used in variation with *tu* 'you.SG' to indicate proximity and intimacy (3) in BP:

(3) Espero que ***Você*** me mande dizer onde pretende residir por enquanto, ou que resolução ***tomas***
'I hope that **You** (2SG.NOM.V) tell me where you want to live for the moment, or which decision **you** (2SG.NOM.T) **have made**'
(Letter written by Carlos to Rui Barbosa, 19th century, in Callou/Barbosa 2011, 65)

This linguistic change spread throughout the community as a whole, and the forms lost their initial semantic value (respect and reverence). During the 20th century, the form of address *o senhor* 'you-formal' (literally, *The Lord*, *Mister*, *Sir*) occupied the place left by *você* in the formal system of address forms.

Figure 1 shows the old pronominal system of address forms that were used until the 18th century in Old Portuguese in different syntactic positions, such as subject (nominative), verbal complements (accusative and dative), and possessive forms (genitive). We can see a **symmetrical paradigm** for less and more formal pronouns of address:

Figure 1: Forms of address in Old Portuguese until the 18th century.

Portuguese	**Until the 18th century**	**NOMINATIVE**	**ACCUSATIVE**	**DATIVE**	**GENITIVE**
2nd person pronouns	**Less formal**	*Tu*	*Te*	*Te, a ti*	*Teu*
	More formal	*Vós*	*Vos*	*A vós*	*Vosso*
		Vossa Mercê	*Vossa Mercê, o/a*	*a/para Vossa Mercê, lhe*	*Seu/ de Vossa Mercê*

This pronominal system of address, however, changes mainly when a new informal or neutral paradigm rises. The construction *Vossa Mercê* suffered phonetic erosion (*Vossa Mercê > vosmicê > você > cê*) and semantic bleaching. High frequency of use caused the loss of the original pragmatic information of politeness and respect. The coexis-

tence of two subsystems *Vossa Mercê* > *você* and *tu* 'you.SG' led to the fusion of both, resulting in the emergence of a suppletive or mixed paradigm with forms inherited from the original paradigm of the 2nd person (*tu* system) and forms taken from the new grammaticalized pronoun *você*. Figure 2 illustrates this new pronominal system:

Figure 2: Forms of neutral address in Modern Brazilian Portuguese: a new mixed pronominal paradigm.

Brazilian Portuguese	**NOMINATIVE**	**ACCUSATIVE**	**DATIVE**	**OBLIQUE**	**GENITIVE**
2nd SG pronoun	*tu*/*você*	*te*/*você*/*lhe*/*o(a)*	*Te*/*lhe*/*a você*/*a ti* *para você*/*para ti*	*prep. + você*/*ti*	*teu*/*seu*

As more recent diachronic research has demonstrated, this hybrid pronominal paradigm has been attested in manuscripts written in BP since the 18th and 19th centuries (Lopes/Cavalcante 2011; Lopes et al 2018; Lopes 2019; Lopes/Marcotulio 2019). In (4), for example, we find a combination of *você* (*Fique você*) with forms of the old (verb)-pronominal paradigm expressed by verbal morphology (*tu deves, satisfizeres*) or by the clitic form *te* 'you.DAT'. In (5), we see a "mixture of forms" with the clitic *te* 'you.ACC', which belongs to the old system, co-occurring with the possessive *sua* 'yours' from the *você*-paradigm:

(4) como ***tu*** *me***deves** resposta dehuã grande Carta que ***te*** escrevi, quando ***Satisfizeres*** esta divida: eu mefarei novamente devedor. Fique **você** embora com ósseo Sigarro em quanto eu cá vou uzando da minha agoá fria

'as **you** (2SG.NOM.T) owe me an answer from the long letter that I wrote to **you** (2SG.DAT.T),[4] when **you** (2SG.NOM.T) pay this debt: I will be in debt again. **Enjoy** (you.2SG.NOM.V) your cigarette while I will enjoy my cold shower'
(Letter written by Marquês do Lavradio to Dom Antonio de Noronha, 18th century, in Marcotulio 2010, 145)

(5) com muitas saudades ***te*** abraça ***sua*** dindinha do coração
'with lots of regards **your** (2SG.GEN.V) beloved godmother hugs **you** (2SG.ACC.T)'[5]
(Letter written by Barbara Ottoni to Mizael, 19th century, in Lopes 2005, 219).

4 DAT = dative.
5 ACC = accusative.

For the subject position, this new 'mixed paradigm' (*tu* and *você*—neutral or informal) has a regional distribution in BP.[6] Since the 18th century, 3 subsystems of address can be identified in subject position (nominative): (i) only *tu* (6), (ii) only *você* (7), and (iii) *você*/ *tu* in variation (8):

(i) only *tu*:

(6) ***tu*** resolverás como ***entenderes***, meu querido anjo
'**you** (2SG.NOM.T) will solve it as **you** (2SG.NOM.T) **understand**, my dear angel' (19th century)

(ii) only *você:*

(7) ***Você*** va pedindo a mamãe que lhe ensine a ler e escrever
'Ask (**you.**2SG.NOM.V) your mom to teach you how to read and write' (19th century)

(iii) você/tu:

(8) aqui se falla muito que ***Você*** está ganhando dinheiro como advogado, que hé muito procurado, que ***tens*** conferencias com os homens mais notaveis
'here people say that **You** (2SG.NOM.V) are earning money as a lawyer, that you are very requested, that **you** (2SG.NOM.T) **meet** the most notable men' (19th century, in Callou/Barbosa 2011, 65)

Besides the converging and diverging aspects observed in the subject position, there has been an irregular and non-symmetrical insertion of *você* in the rest of the second person paradigm. Regarding the accusative, the clitic *te* 'you.ACC' has been the original second person direct complement ever since Latin. Nonetheless, the implementation of *você* into the pronominal system brought the forms of the third person paradigm to the second person: the strong pronoun *você* and the clitics *lhe* 'you.3SG' in addition to *o*/*a* 'him/her = you' already in use in EP for direct objects corresponding to *você*. Studies based on private letters written by Brazilians from the 18th and 19th centuries identify the predominance of clitic *te* 'you.ACC', proclitic to the verb, mainly in Southeastern and Southern regions (cf. Lopes/Marcotulio 2019; Lopes et al 2018; Souza/Oliveira 2013; Coelho/Gorski 2011; Nunes/Coelho 2013). However, in the Northeastern region, the forms of the *você* paradigm prevailed: the clitic *lhe* 'you' in Bahia letters and *o*/*a* in Pernambuco letters (cf. Lopes/Rumeu/Carneiro 2013; Gomes/Lopes 2014).

6 In some regions, predominantly in the central area of Brazil, the entire Southeastern area (except for Rio de Janeiro, where the subsystem *você*/*tu* prevails), and some Northeastern regions, the frequency of *você* is higher or even exclusive. However, in other regions, such as the South, the pronoun *tu* is most frequent. It is noteworthy that we can find variation between the two pronouns (*você* and *tu*) in all regions of Brazil, except for some in the South (cf. Scherre et al. 2015, 133–172).

Besides its predominance as an accusative complement, the clitic *te* 'you' also prevailed as a second person dative in proclitic position. In the remaining documentation of the 19th and 20th centuries, the clitic *te* was very frequent in the Southeastern (Rio de Janeiro and São Paulo) and Southern (Santa Catarina) areas of Brazil (Nunes/Coelho 2013). The use of the clitic *lhe* is more frequent in the Northeastern areas (Rio Grande do Norte, Pernambuco and Bahia) (cf. Martins/Moura 2013; Gomes/Lopes 2014; Lopes/Rumeu/Carneiro 2013).

Other dative variants were identified as well: prepositional phrases associated with the *tu* paradigm (*a*/*para ti* 'to/for you') were more frequent in the Southern areas, whereas the prepositional phrases associated with the *você* paradigm (*a*/*para você* 'to/for you') were more frequent in the Northeastern region.

As for oblique relations, where pronouns function as complements of prepositions, the forms of the *você* paradigm presented high levels of use in most parts of the Northeast and Southeast (Minas Gerais). In Rio de Janeiro (Southeast) and Santa Catarina (South), however, prepositional phrases with *ti* are more common.

The variation in the genitive forms presented the same behavior observed in oblique forms. The form arising from the *você* paradigm—the older third person possessive (*seu* 'your')—prevailed over the original second person possessive (*teu* 'your') in letters written in the same regions.

In summary, we observe that the most favorable contexts for the resistance of *tu* forms are the accusative and dative complements. The forms of the *você* paradigm, as ancient third person forms, are more easily found in other syntactic contexts such as the oblique and genitive. In the following sections, we shall show that the reorganization of the pronominal paradigm had important corollaries in 2 areas of syntax—the syntax of clitics and the broader use of null objects.

2.2 The new syntax of clitics

In correlation with the facts presented above, a new syntax of clitics emerged in BP. First, it is important to stress that the second-person clitic pronoun *te* 'you.ACC.DAT' entered in free variation with the non-clitic pronoun *você* 'you',[7] as can be seen in (9) and (10):

(9) Eu no sábado espero **você** (20th century)
I on Saturday wait **you** (2SG.ACC.V)
'On Saturday, I will wait for you'

7 This is also true for the first-person clitic pronoun *me* 'me', which varies with the nominative pronoun *eu* 'I' in some dialects of BP. Since this variation is much more widespread in the second person, this is what we shall focus on, leaving aside the first person. For further details and analysis, see Galves 2018, 2019.

(10) **Te** acompanharei em muitas orações (20th century)
You (2SG.ACC.T) will accompany in many prayers
'I will accompany you in many prayers'

This state of affairs is original within Romance varieties, where tonic pronouns normally do not appear in direct object position, as in (9). This innovation in BP is associated with another novelty concerning the placement of clitics. Different from what is observed in other Romance languages, including EP, in verbal locutions, clitics in BP normally affix to the verb which they are the arguments of rather than raising to the finite auxiliaries, as can be seen in (11) and (12), in which the verb is in the past participle form and in the infinitive form, respectively. In these sentences, the presence of adverbial expressions between the tensed verb or auxiliary and the non-tensed form of the thematic verb evidences the fact that the clitic is affixed to the latter and not the former:

(11) porque em toda sua vida **tem** por artimanhas **se apossado** de uma
because in all his life (he) **has** through deceit **CL-REF seized** of a

grande parte da fortuna
great part of+the fortune
'because, during his life, he seized great part of his fortune through deceit'

(12) **Quero** ainda uma vez **lhe agradecer** as boas e
(I) want still one time **CL-DAT thank** the good and
generosas| palavras (Letter, in Carneiro 2005)
generous words
'Once again, I want to thank you for your good and generous words'

Additionally, BP innovates with respect to the position of clitics in sentences in which the verb is in absolute first position. Throughout the history of Portuguese, this had been a context in which enclisis was obligatory. In 19th century texts written in Brazil, we find several occurrences of pre-verbal clitics in this context:

(13) **Me** **entendi** com o João Victorino a serca do seo boi
Me (1SG-ACC) **understood** with the João Victorino about of+the his ox

que matarão (Letter 418, in Carneiro 2005, 146)
that (they) killed
'João Victorino and I reached an agreement about his ox that was killed'

The striking fact about the diachrony of clitic placement in BP is the emergence of variation between enclisis and proclisis in all syntactic contexts. This stands in contrast

with BPs ancestor, Classical Portuguese, in which variation was limited to well-defined contexts (cf. Galves/Britto/Paixão de Sousa 2005), with either enclisis or proclisis being categorical in the others. For instance, we find variation in contexts that in the past were exclusively proclitic such as sentential negation or relative clauses. Examples (14) and (15) illustrate the occurrence of enclisis in those contexts:

(14) O portador vai com ordem de puchar, **náo demore-o**
the bearer goes with order of pulling, **not delay him**
'The bearer (of the message) has been ordered to hurry, do not delay him'

(15) ... uma depressão na região precordial, **que durou-me** até
a depression in+the region pre-cordial **that lasted me** until
as 5 horas da tarde. (Letter 132, in Carneiro 2005, 164)
the 5 hours of+the afternoon
'... a chest tightness, which affected me until 5 o'clock'

The across-the-board enclisis/proclisis variation in the attested history of BP derives from two opposing tendencies. One can be considered the natural trend of BP—the generalization of proclisis; the other is due to the strong influence of EP, in which enclisis had become the general pattern in the former enclisis/proclisis variation (cf. Galves/Britto/Paixão de Sousa 2005).[8] This influence can be seen in Figure 3, from Carneiro/Galves (2010), in which the three lines correspond to the frequency of enclisis in variation contexts in three different corpora. The triangles are from Galves/Britto/Paixão de Sousa (2005) and show that a change in placement that favors enclisis takes place in Portugal around 1700 and is complete around 1850. The values represented by squares and diamonds are from Pagotto (1992) and Carneiro (2005) respectively. The first period in Pagotto's corpus corresponds to Classical Portuguese. Carneiro's corpus is entirely from the 19th century. Both studies show that the frequency of enclisis in Brazilian texts significantly increases from the beginning of the 19th century to the beginning of the 20th century. Both corpora also evidence the inversion of this tendency from 1920 on.

8 In Classical Portuguese, variation contexts are contexts in which the verb is not in first position (which would require enclisis) and is not preceded by any elements forcing proclisis, like negation, subordinating conjunctions, interrogatives, focus words, or certain adverbs.

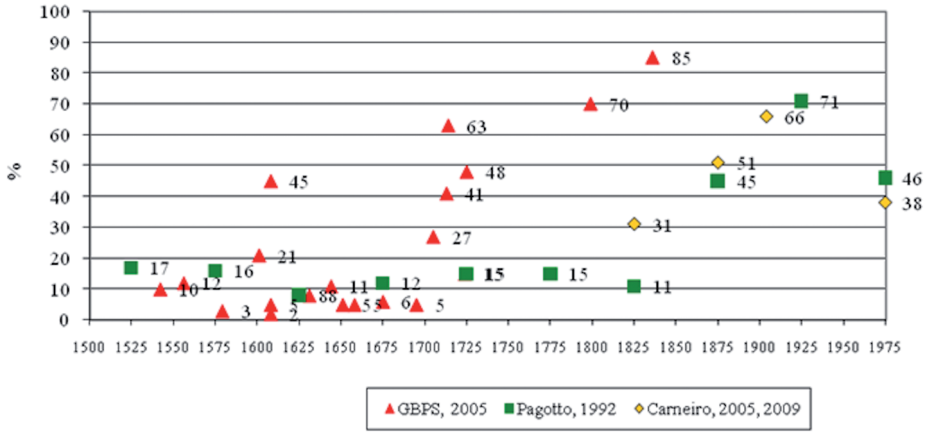


Figure 3: The evolution of enclisis in non-dependent V2 clauses (Galves/Britto/Paixão de Sousa 2005, sorted by the birthdate of the authors; Pagotto 1992 and Carneiro 2005, sorted by the date the texts were written).

Both the parallelism with the evolution of European Portuguese and the failure of the change during the 20th century suggest that enclisis in Brazilian texts of the 19th century is the effect of the prestige of European Portuguese at that time. This is not incompatible with the claim that the natural tendency of the Brazilian vernacular is proclisis, even if, for a long time, this is blurred in texts due to the effect of the Portuguese norm. From the point of view of the development of a new vernacular pattern, however, it is possible to say that BP underwent two changes. First, the Tobler-Mussafia Law, which prevents clitics from appearing at the beginning of Intonational Phrases, is lost, as shown by sentences like (13). Second, clitics cease to be elements associated with Inflection (Infl clitics) to become affixes of the verb (V-clitics), as can be seen in (11) and (12) (cf. Galves/Ribeiro/Torres Moraes 2005; Lopes/Souza/Oliveira 2013).

In the next section, we shall see another phenomenon related to the emergence of a new pronominal syntax in BP, the loss of the third person clitic, and its replacement by null objects in use.

2.3 Deletion of the 3rd person accusative clitic *o*/*a* 'him/her'

While first- and second-person clitics are still present in vernacular BP, the third person clitic has virtually disappeared from speech. Studies show that it is not acquired by children before school (Correa 1991) and that the preferred strategy for the anapho-

ric 3rd person is the zero form (Duarte 1986).[9] In her diachronic study on theatre plays, Cyrino found a strong decrease of explicit pronouns in object position, from 90 % in the 16th century to less than 20 % in the 20th century (Cyrino 1997, 246). As for null arguments with specific antecedents, their frequency goes from 3 % to 67 % during the same period. In a corpus of Brazilian newspapers published in the middle and the end of the 19th century and in the middle of the 20th century, respectively, Macedo Costa (2012) observes a similar, although less drastic, phenomenon.

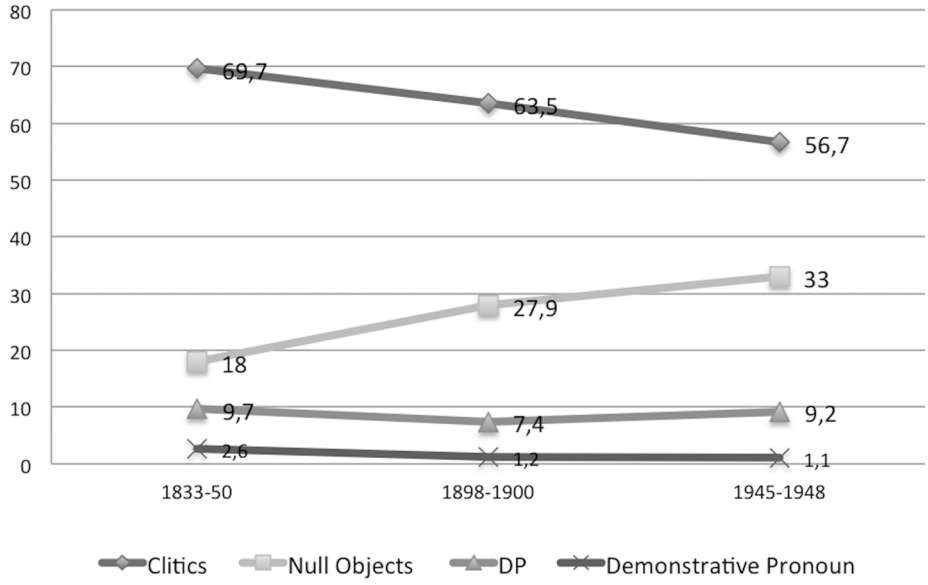


Figure 4: The evolution of the frequency of the four types of BP anaphoric direct objects (1st phase: 1833–1850; 2nd phase: 1898–1900; 3rd phase: 1945–1948) (Macedo Costa 2012, 166).

Figure 4 shows that the frequency of clitics decreases from 70 % to 57 % while the frequency of null objects increases from 18 % to 33 %. In other words, the difference in frequency between them is significantly reduced, from 50 to 23 percentage points. According to the same study, the frequency of null objects with a specific nominal expression as their antecedent increases from 10 % in the first phase to 25 % in the third phase.

In conclusion, both Cyrino's and Macedo Costa's studies evidence a change over time in the expression of anaphoric objects, as expected. The quantitative difference

9 The other strategy for avoiding clitics in speech is the use of the tonic pronoun *ele* 'he', in parallel with what occurs with the second-person pronoun *você* 'you'. But this use is stigmatized and consequently is avoided in written texts, which makes it difficult to study its introduction into the system. It must be observed, however, that it appears three times in the *Proceedings of the Sociedade Protetora dos Desvalidos* (Oliveira 2006), cf. Galves e Lobo (2019).

observed between newspapers and plays is very likely derived from the difference between the formal written language used in the former and the familiar spoken language represented by the latter.

In his pioneer study, Tarallo (1983) points out that while in object position, the tendency over time is for pronouns to be null, the reverse happens in subject position, where the frequency of lexical subjects increases from 23.3 % in the 18th century to 65.6 % in 1982. In the next section, we look in deeper detail at the evolution of null subjects in the history of BP.

2.4 Null subjects

Both Classical Portuguese and Modern European Portuguese are what has been called "consistent" null subject languages in the recent literature. This means that there are no syntactic restrictions on the distribution of null subjects, and their interpretation is essentially the same as that of personal pronouns. In this kind of language, the null subject-personal pronoun variation is mainly due to conditions imposed by the information structure of texts. By contrast, many researchers have long noticed that null subjects are not allowed in all contexts in BP (cf. the studies in Kato/Negrão 2000) and that third person singular null subjects, in the absence of a possible binder, can be interpreted as indeterminate. This is impossible in European Portuguese, in which either the 3rd person plural or the adjunction of the clitic *se* is required to obtain this interpretation (Galves 2001, [1]1987). Thus, while the preferred BP interpretation of (16) is that people in general ceased to use horse-bits, it only means that a specific person ceased to use horse-bits in EP, and the generic interpretation requires the presence of *se*, as in (17):

(15) Não usa mais freio
NEG use-3SG more horse-bit

(16) Não se usa mais freio
NEG CL-REF use-3SG more horse-bit
'One no longer uses horse-bits'

Occurrences of indeterminate null subjects without *se* are not frequent in historical records. Some cases were found in the first half of the 19th century in advertisements (cf. ex. 18 from Avelar/Cyrino 2014). They are also observed in the proceedings of the *Sociedade Protetora dos Desvalidos*, written by ex-slaves in Bahia (cf. ex. 19, drawn from Klebson Oliveira's thesis):[10]

10 For more details on the syntax of clitics in these documents, see Galves/Lobo (2019)

(17) Vende huma Negra Cozinheira (Pernambuco 1829)
sell.3SG a black.FEM cook.FEM
'Black cook for sale'

(18) E se não for bom reúne
and if NEG be.FUT.SUBJ.3SG good meet.3SG
'And if things don't do well, people meet'

In accordance with the evolution of the analysis of the Pro-Drop Parameter,[11] many researchers now agree that instead of displaying a change in progress from a null-subject to a non-null-subject language, BP underwent a change from a consistent to a partial null-subject language.

Duarte (2012) studied the evolution of the occurrences of personal pronouns in a corpus of Brazilian plays written from 1845 to 1992. Figure 5 shows the evolution of the use of lexical pronouns according to person:

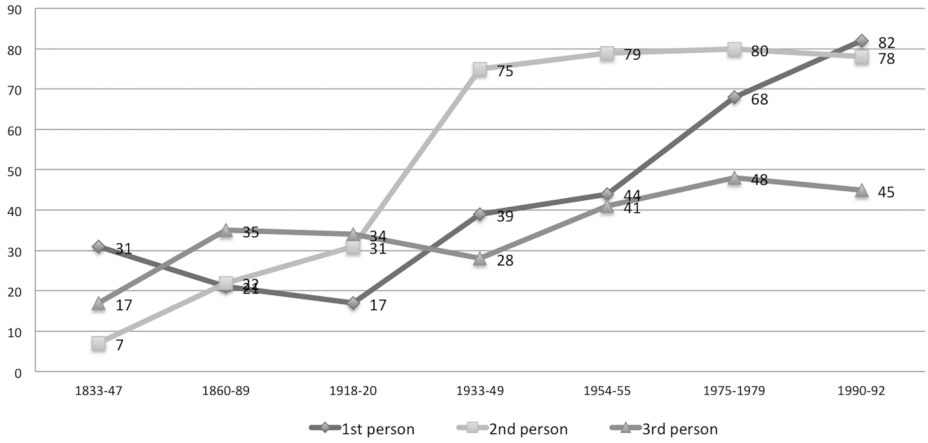


Figure 5: The evolution of the use of lexical pronouns according to person (Duarte 2012, 22).

This graph shows that, in addition to the general increase in use of lexical pronouns, the second person displays different diachronic behavior, having undergone a sharp rise between the third and fourth period (1918 and 1937–1938), which is, according to the author, due to the substitution of *você* for *tu* 'you'. The first person initially accompanies the tendency at a slower pace, but ends up undergoing a strong increase in frequency from the fifth period (1955) on, reaching the same frequency as the second person in the last period (1990–1992). The dynamics of the third person, however, are different. As already been pointed out by Duarte (1995), the third person is where null

11 For a recent overview, see Biberauer et al. (2010).

subjects are still more frequent nowadays. Moreover, we observe an initial shift in frequency of the lexical pronouns, from 17 % in the first period (1845) to 35 % in the second period (1882–1883). This fact is striking for two reasons. First, the rate of 17 % of lexical pronouns is close to what we observe in Classical Portuguese at the same time (data from the *Tycho Brahe Corpus*, cf. Carpani 2010). Second, another diachronic study based on data collected from newspapers evidences a change between the middle and end of the 19th century, with frequencies similar to what we observe in the two first periods of Duarte's study, as can be seen in Figure 6 (Gravina 2008, 89):

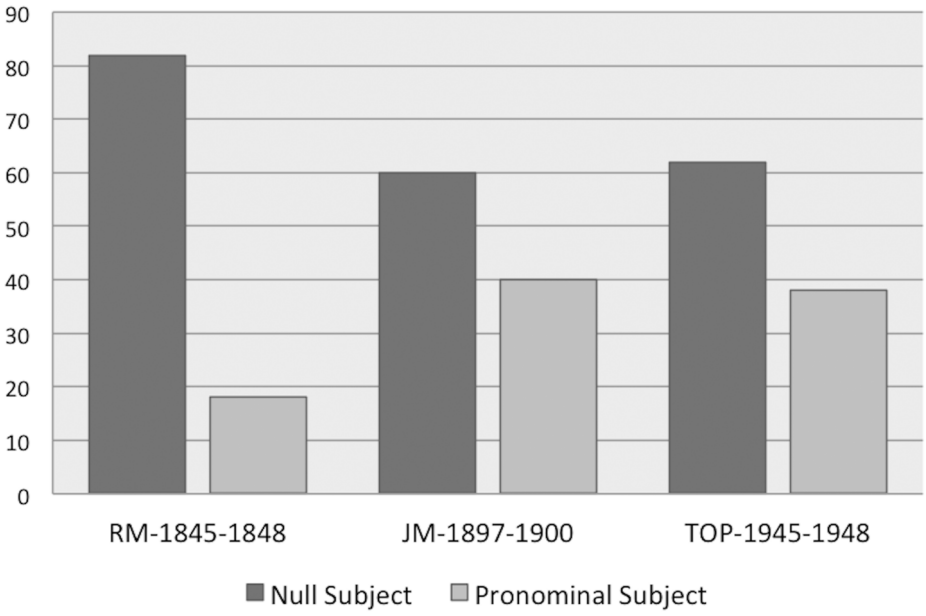


Figure 6: Frequency of null subjects (vs. pronominal subjects) in newspapers in three time periods.[12]

The claim, based on the quantitative data above, that a new grammar of null subjects emerged during the second half of the 19th century is reinforced by more qualitative evidence provided by the same corpora. First, Duarte (2012, 33) observes that in the first period (1845), no lexical pronouns appear that refer to an antecedent in subject position in the same sentence and only in 6 % of the cases when the antecedent is the subject or topic of an adjacent sentence. This is what we expect in a consistent null subject language, in which lexical pronouns are typically contrastive or emphatic. In the second period, however, lexical pronouns begin to appear in those contexts at a

12 RM = *Recreador Mineiro* newspaper; JM = *Jornal Mineiro* newspaper; TOP = *Tribuna de Ouro Preto* newspaper.

rate of 6 % and 18 % respectively, indicating lexical pronouns' losing of their contrastive function. In (20), drawn from a corpus of data taken from advertisements, we observe that cases of lexical pronouns whose antecedent is in the matrix clause were already occurring in the first half of the 19th century (Torres Moraes 2002):

(19) João Francisco$_i$ ..., aviza ao Respeitavel Público d'esta Cidade
João Francisco announces to+the respectable Public of+this town

que elle$_i$ tem um sortimento de espingardas de um novo systema
that he has a selection of rifles of a new system
(São Paulo, 1828)

Another significant difference is found by Gravina (2008) between the first and second periods. In the former, the rate of null subjects in relative clauses and in interrogative clauses is comparable to the mean frequency of null subjects, that is, around 80 %, which means that lexical pronouns are as disfavored in those contexts as in all others. In the second period, however, she finds more lexical pronouns (57 % of the cases) than null subjects (43 %) in these types of clauses. This indicates that this context has begun to be disfavored, since the overall rate of null subjects (60 %) is still higher than the rate of lexical pronouns. This is the tendency that we expect if there is a change towards a partial null-subject language. In effect, null subjects in such languages are no longer licensed by a strong inflection on the verb but by a binding relation with some antecedent in subject or topic position (cf. Kato/Negrão 2000). The clauses introduced by a WH-operator, such as relative, interrogative, and other clauses, act as syntactic islands that block such a binding relation between the null subject and its antecedent.

In the next section, we shall address the question of the nature of the changes discussed so far.

2.5 The abstract nature of the changes in the pronominal morpho-syntax of Brazilian Portuguese

There is a consensus among researchers that the fact that BP null subjects behave differently from those of other Romance pro-drop languages is correlated with the weakening of its verbal morphology. It must be emphasized, however, that BP did not evolve towards a non-null-subject grammar (like French, for instance); it became a partial null-subject language. In some analyses, this development is explicitly correlated with the topic-oriented properties of the language, which we shall return to below (cf. Section 4.1). Going a step further, Negrão/Viotti (2000) argue that BP's restrictions on the use of null subjects do not derive from a weakening of the inflectional morphology but are instead a consequence of the discourse-oriented properties of the

language, which entail that "the recoverability of empty categories occurs via discourse prominence" (Negrão/Viotti 2000, 109).

It cannot be disputed, however, that there is a relationship between the weakening of the verbal morphology and the use of new pronouns, which is bidirectional. The substitution of the 2nd person pronoun *tu* 'you' by the pronoun *você* 'you' (a former form of address *Vossa Mercê* 'Your Grace' with the verb in the third person singular—cf. Section 2.1) in many dialects had the effect of reducing the paradigm from five to four forms. In these conditions, the third person affix on the verb becomes ambiguous with respect to whether it refers to the second or the third person of the discourse. The variation of the first person plural pronoun *nós* 'we' with the former nominal phrase *a gente* (currently *we* but originally *the people, one*) is frequently mentioned as a further source of reduction in conjugations. These pronominal paradigms are summarized in Figure 7:

Figure 7: Pronominal paradigms for subject position

	Old paradigm		**New paradigm**	
	Old Pronouns	**Verb *falar*** 'to speak'	**Emergent Pronouns**	**Verb *falar*** 'to speak'
1SG	**Eu**	fal-**o**	Eu	fal-**o**
2SG	**Tu**	fala-**s**	Tu[i]/**Você**	}falaØ
3SG	**Ele/ela**	falaØ	Ele/ela	
1PL	**Nós**	fala-**mos**	Nós/ **A gente**	fala-mos falaØ[ii]
2PL	**Vós**	fala-**is**	**Vocês**	}fala-**m**
3PL	**Eles/elas**	fala-**m**	Eles/elas	

Obs[i]: In some dialects of BP, the verbal ending is present.
Obs[ii]: In 1PL, the agreement is also variable.

Crucially, however, the weakening of the verbal paradigm cannot be completely reduced to the change in the use of pronouns. It also appears in the lack of agreement in person and/or number between pronominal subjects and verbs. A clear example of this fact is that in most dialects in which it is still in use, the second person pronoun *tu* 'you' does not trigger the second person affix –s on the verb (*tu sabe* instead of *tu sabes* 'you know'). The erosion of morphology within the verbal paradigm, which is also visible in the nominal paradigm, is likely to be an effect of the intensive process of linguistic contact in Brazil (↗2 The social history of Brazilian Portuguese, ↗3 The history of linguistic contact underlying Brazilian Portuguese, ↗5 Historical morphology). But it must be emphasized that not all inflectional morphology has been lost and that, in particular, all Brazilian dialects have kept the morphological distinctions be-

tween the first and third person markings on verbs that already existed in European Portuguese. Galves (1993) argued that such distinctions could be interpreted as the expression of a binary feature +/- Person. In other terms, Person in BP does not have the three values defined discursively but only a positive or a negative value, just the same as Number. According to Galves, this would characterize weak Inflection,[13] incapable of identifying null subjects.

In Section 4 below, we return to Brazilian innovations related to subjects and topics, which can be explained by a change in the features of the inflectional paradigm.

3 Morpho-syntax and the use of verbal constructions

3.1 The Periphrastic Future

The replacement of simple forms with periphrastic forms lies at the origin of future expressions formed in Portuguese and several other Western Romance languages. From the Latin SCRIBERE HABET (to-write s/he-has) 's/he has to write', the following originated: (il/elle) *écrira* in French, *scriverà* in Italian, *escribirá* in Spanish, and *escreverá* in Portuguese 's/he will write' (Clackson/Horrocks 2011, 279). This phenomenon involved a loss of morphosyntactic autonomy in the auxiliary verb HABERE and a strong phonetic reduction in Romance languages. The new Romance future form became synthetic and thus no longer periphrastic as it was in Latin.

In Portuguese, a new cycle started with the grammaticalization of the verb *ir* 'to go' as a future auxiliary in constructions of the type '*ir* +INF' (e.g., *Ele vai terminar o curso no próximo mês* 'He is going to finish the course next month'). The grammaticalization of movement verbs such as *ir* is not restricted to Portuguese and occurs in several other languages such as English (*be going to*), French (*aller* + INF), and Spanish (*ir* + *a* + INF).

Essentially, full verbs that embody the idea of "movement in physical space" work as auxiliary verbs indicating "movement into future time" in periphrastic structures. Since displacement in physical space takes some time, the meaning of the periphrastic verb *ir* is eventually extended to modality, the fulfilment of a wish, or the projection of an action to a later time until the verb is reanalyzed as a future auxiliary.

In Old Portuguese, context which favored change would be formed by the verb *ir* combined with a non-finite clause indicating purpose, as in (21):

---

13 In generative grammar, Inflection (Infl) is considered a functional category that heads clauses. In the older versions of the theory, Infl was composed of the features Agreement (Agr) and Tense (T). In the 1980s, such features came to be considered categories of their own. More recently, Chomsky (1995) argued that Agr had to be excluded from the Functional Categories because of its lack of interpretation at the semantic interface LF.

(20) Recolhee vosa jemte que vem fogindo e **vamos** **a** pelejar.
gather.2PL your people who come escaping and go.1PL to fight
'Gather your people who are escaping and we go to fight'
(*Crónica de Portugal*, 15th century)

In (21), we can still identify two clauses: [vamos 'go'] and [a pelejar 'to fight']. The verb *ir* (*vamos* 'we go') would act as a predicative verb in its original basic value of displacement in space. The next clause [a pelejar 'to fight'], starting with the preposition *a*, introduces the aim—the intention to be fulfilled—projecting the action to be performed towards a near future. The corresponding interpretation would be: 'we will move in space (in order to) fight'. We should stress that the preposition was not required in Old Portuguese, as can be observed in (22):

(21) Soomemte vos direy o que pemso de fazer: nos **vamos** tomar
only you tell.1SG what think.1SG of do.INF we will take

nossas çilladas de noite
our ambushes at night
'I only tell you what I think to do: we will take our ambushes at night'
(*Crónica do Conde D. Pedro de Meneses*, 15th century)

Josane Oliveira suggests that the grammaticalization process of *ir* as a future auxiliary in Portuguese started slowly in the 16th century, evidenced the fact that the ambiguous interpretation was recurrent in the data analyzed in the study. In (23), two readings are possible. On one hand, the spatial movement performed by the human subject is expressed by the first person of the verb *ir* (*yrey* 'go-1P-FUT'), and the purpose of this displacement appears in the infinitive form (*morar alá* 'live there'). The intervening element (*contigo* 'with you') supports this analysis. On the other hand, the future interpretation is also possible (*yrey morar* 'I will live'), enabled by the morphological future marker in the auxiliary verb *ir* (*yr<u>ey</u>* '*go*-1P-FUT'):

(23) eu me *yrey* contigo *morar* alá
I me will-go with you live there
'I will go to live with you there'
(16th century, in Oliveira 2006, 94)

The study analyzed the representation of changing future forms[14] in documents written in Portuguese between the 18th and 20th centuries. For this time period, only docu-

14 The following were controlled: simple or synthetic future (*cantarei* 'I will sing'), periphrasis with *haver de* 'shall' in the present tense and in future + infinitive (*hei*/*haverei de* cantar) constructions, peri-

ments in BP from the corpus of the project PHPB-RJ[15] were considered. The chart below shows the behavior of the synthetic form as opposed to the periphrastic structure:

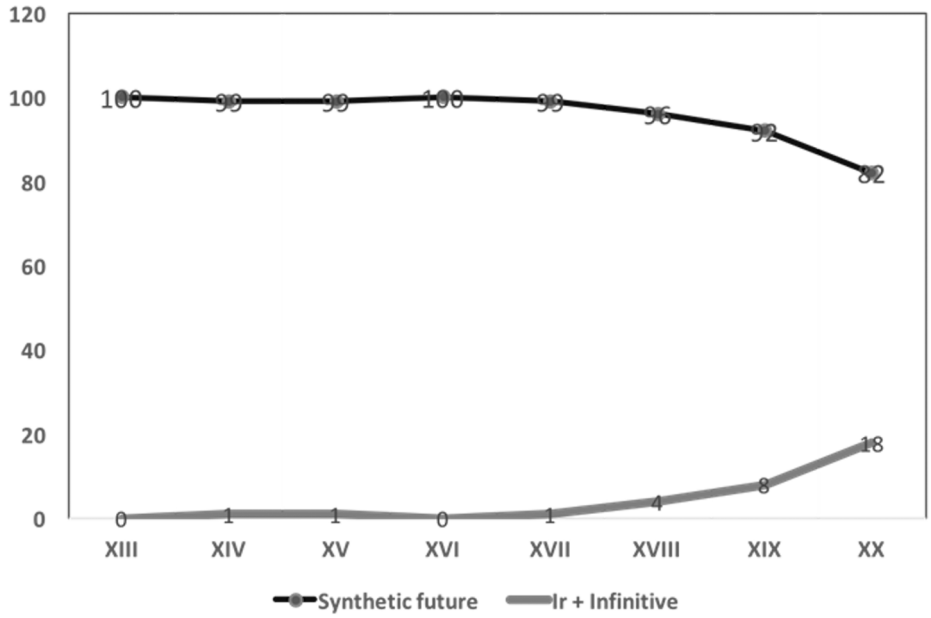


Figure 8: The evolution of the frequency of Ir +Infinitive in Portuguese written texts adapted from Oliveira (2006, 92).

Regarding only the variation between the simple future and the periphrastic construction (*ir*+INF), analysis revealed that the simple future presented high frequency levels throughout the observed period. The frequency of the periphrasis (*ir*+INF) was below 2 % between the 14th and 17th centuries. However, from the 18th century on, the frequency of usage of the periphrasis with *ir* progressively increased in texts written in Brazil to 4 % in the 18th, 8 % in the 19th, and 18 % in the 20th century.

In sum, the author argues that, in Portuguese, auxiliarization in the construction (*ir*+INF) started in the 16th century and gradually increased in frequency. From the 18th century on, periphrasis starts to become more prolific in written texts. The synthetic future form prevailed in writing up to the 20th century. Regarding speech analysis, the author suggested that the replacement of the synthetic future with the periphrastic form with *ir* had become evident by the end of the 20th century. In this modality,

phrasis with the verb *ir* in present and future tense (*vou*/*irei* cantar 'I will sing'), and the present tense (*canto* 'I sing').

15 The part of the *History of Brazilian Portuguese Project* developed at the Federal University of Rio de Janeiro.

mainly in BP, the periphrastic construction spread to new contexts such as stative verbs, which would be ungrammatical with the verb *ir* in its original usage:

Vou ficar com você 'I will stay with you'.
Vou estar enviando 'I will be sending'.
?*Vou ir contigo* 'I will go with you'.

The first two cases do not create a negative reaction for BP speakers. The last, however, is still avoided in the speech of particular social groups, different than observed in Spanish, for instance, with *Voy a ir*.

3.2 *Haver* and *ter* 'have' in existential constructions

Synchronic and diachronic studies suggest that the variation and change process in existential constructions with the verbs *haver* and *ter* 'have' is another morpho-syntactic aspect which strongly distinguishes BP from EP. While in BP, existential *ter* is preferred; in EP, *haver* almost categorically indicates existence:

(24) EP/BP: **Há** algumas pessoas aqui.
exist some people here
'There are some people here'

(25) BP: **Tem** pessoas aqui.
have people here
'There are people here'

The variation between *haver* and *ter* in existential constructions in Portuguese exhibits several overlapping layers of change. One such example exists between the verbs *ser* 'be' and *haver* 'have', when the former used to work as an existential in old Portuguese in constructions such as (26), whereas *haver* only occasionally expressed that value, as in (27):

(26) Aniball, que **foy** no tempo dos gemtios
Aniball, that **was** in+the times of+the Gentiles/Pagans
'Aniball who existed in the times of the Gentiles'
(*Crônica de D. Pedro*, 15th century, in Brocardo 2014, 129)

(27) nom **houve** no paaço quem podesse entender per u
not **there was** in+the palace who could understand for where

Galaaz entrara
Galaaz entered
'There was not anyone in the palace who could understand how Galaaz got in'
(*Demanda do Santo Graal*, 15th century)

The other layer of change is found in possessive constructions, in the variation between *haver* and *ter*, which carried values related to possession in their Latin meaning: "possess, hold" for *habere* and "have in one's hand, obtain" for *tenere*. In texts in Old Portuguese, *haver* was more productive than *ter* with all possession values: [+material], as in (28) and [-material], in (29). Nevertheless, from the 15th and 16th centuries onward, the verb *ter* replaces *haver* in all kinds of possession, (in 30 and 31), whereas *haver* specialized as an existential verb, as in (33), eliminating the etymological *ser* (Mattos e Silva 1997):

(28) Senhor, ora **hei** já a espada, [...] o escudo nom **hei [+material]**
Lord, now I-have yet the sword, the shield not I-have.
'Lord, now I've got the sword, [...] I don't have the shield'

(29) E os que **ouverõ** nome dom Fernando **[-material]**
and those who **had** name Dom Fernando
'And those who were called Sir Fernando'

(30) quen quer que **tenia** meu tesouro **[+material]**
whoever that had my treasure
'whoever had my treasure' (in Mattos e Silva 1997, 270)

(31) E aquele que foge nom deve **teer** dereitura nem lealdade **[-material]**
and the one that escapes not should have justice nor loyalty
'And the one who escapes should not have either justice nor loyalty'
(*Vida e feitos de Júlio Cesar*, 15th century)

(32) em seu tempo neste reino não **havia** mouros que conquistar
in its time in+this kingdom not exist moors that conquer
'At that time there were no moors to be conquered in this kingdom'
(*Décadas de Asia*, 16th century)

The diffusion of *ter* rather than *haver* was not restricted to contexts of possession, since the verb *ter* also gained an existential meaning later on, yet without losing its possessive value in some biargumental contexts of the type X has Y:

Paulo (=X) tinha a carta (=Y)
'Paulo (=X) had the letter (=Y)'

Avelar/Callou (2007) and Avelar (2009) have suggested, based on written documentation from the 17th and 19th centuries, that some constructions with the verb *ter* were ambiguous, allowing both existential and possessive readings. One context which favors an existential reading of the verb *ter* is created by inanimate subjects indicating place names, as in (33). We can observe that the noun phrase (NP) *A Santa Caza de Misericórdia* is a place name, which as such, can be interpreted as the place where such *desconueniencias* 'inconveniences' exist. Likewise, this determiner phrase (DP), as a more abstract possessor, can function as a grammatical subject of *ter*, with the constituent *desconueniencias* being the possession.

(33) deve **a Sancta Caza da Mizericordia** [...] ver as desconueniencias
should the Santa Caza da Mizericordia [...] see the inconveniences

que **tinha** (17th century, in Avelar/Callou 2007, 385)
that **had**

Possessive reading: 'The Sancta Caza da Mizericordia had inconveniences.'
Existential reading: 'There were inconveniences in the Santa Caza da Mizericordia.'

According to Avelar/Callou (2007), the existential *ter*, despite still being ambiguous at that time, progressively spread in BP. In the 18th century, the frequency with an existential value was quite low (8 %); however, between the 19th and 20th centuries, the existential *ter* rose from 22 % to 30 % in the documentation analyzed by the authors.

The difference between BP and EP concerning the variation between *ter* and *haver* in existential constructions began to surface in the 19th century and was consolidated in the next century. While existential *ter* progressively gained ground in Brazil, existential *haver* remained productive in EP. In a study based on stage plays, Marins (2013, 79) reveals, for instance, that until the second half of the 19th century, the use of existential *haver* was equivalent in the Brazilian and Portuguese plays that were analyzed—usage frequency hovered around 90 % for both BP and EP. From the second half of the 20th century on, the productivity of *haver* in BP decreases drastically, whereas in EP data, the frequency levels remain steady throughout the 20th century.

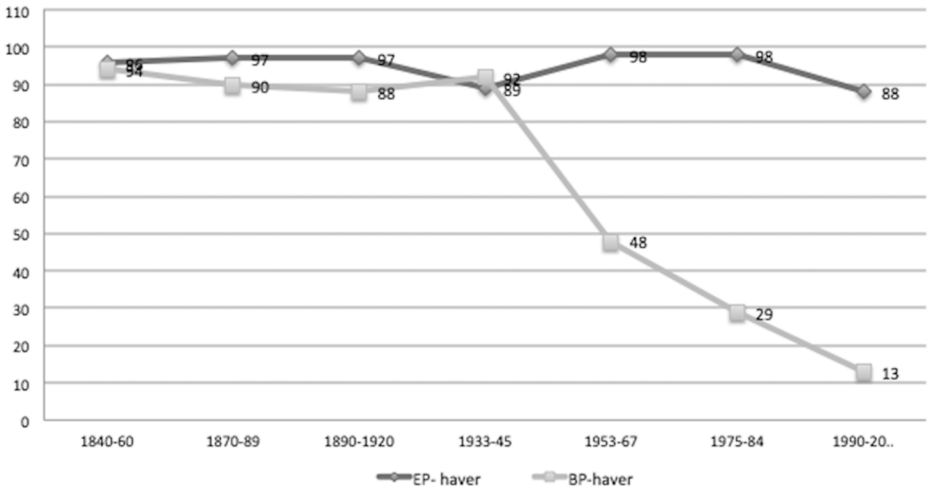


Figure 9: Existential sentences with the verbs haver and ter in BP and EP (adapted from Marins 2013, 79).

Some authors (Avelar/Callou 2007, 386; Avelar 2009, 160) argue that the identified differences are not restricted to usage frequency but instead stand for two grammars with separate courses in both territories. For Avelar (2009), the reanalysis of *ter* as an existential verb is related to other linguistic changes which occurred in BP such as the loss of referential null subjects:

(34) Tem vários móveis velhos na casa.
have much furniture old in+the house.

BP: '**There is** a lot of old furniture in the house.'
EP: '**He/she** has a lot of old furniture in the house.' (Avelar 2009, 160)

This utterance would be interpreted as existential in BP, while in EP, the same sentence would have a possessive reading with a third-person referential null subject. A European Portuguese speaker might identify a possessing subject, even if phonetically empty, for the verb *ter*. The corresponding reading would be 'Someone (he or she) has/owns old furniture'. Given that, the frequency of the referential null subject became strongly restricted in BP, as discussed above, a Brazilian speaker has some difficulty interpreting constructions of the type in (34), without any DP expressed in the pre-verbal or subject position, as being possessive. In these contexts, the verb *ter* would have an existential status reinforced by the locative adverb *na casa* 'in the house' and is interpreted as 'There is a lot of old furniture in the house'. Avelar (2009) argues that a BP speaker does not interpret the example (34) as possessive because that reading would require a null subject, which is no longer licensed in BP grammar.

4 Aspects of the syntax of clauses

4.1 Subjects and topics

The decrease in the frequency of post-verbal subjects (VS) is one of the most analyzed phenomena in BP (Torres Moraes 1993; Berlinck 2000; Gravina 2014; amongst others). Interestingly, a similar change is seen in the history of European Portuguese (Paixão de Sousa 2004; Galves/Paixão de Sousa 2017; Galves/Gibrail, 2018), beginning at roughly the same time. Table 1, from Berlinck (2000), and Figure 10, from Galves/Paixão de Sousa (2017), illustrate this evolution in the two varieties respectively:

Table 1: Word order frequencies in BP across six periods of time (adapted from Berlinck 2000, 183).

Period of time	1730	1780	1830–1850	1880	1930	1970–1990
Word order						
SV	42 %	76 %	90 %	89 %	94 %	96 %
VSX	34 %	14 %	6 %	5 %	2.5 %	2 %
VXS	24 %	10 %	4 %	6 %	3,5 %	2 %

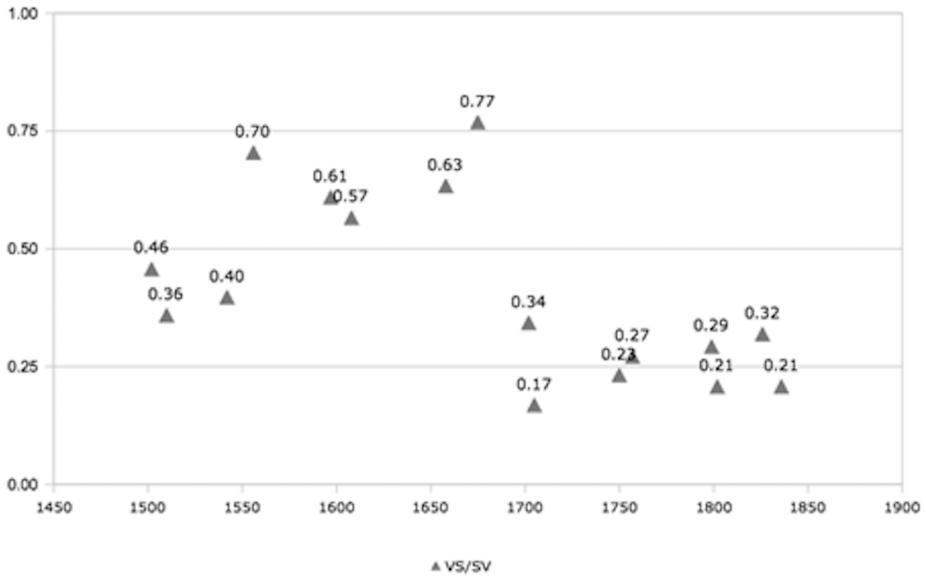


Figure 10: : The evolution of Subject-Verb-Object order in matrix declarative transitive sentences (by author according to birthdate) in the diachrony of European Portuguese.

Antonelli (2011), Galves/Paixão de Sousa (2017), and Galves/Gibrail (2018), among others, argue that the high frequency of VS in Classical Portuguese matrix declarative clauses is due to the movement of the verb to the head of the highest layer of the clause: C.[16] The sharp decrease in this word order observed in the Portuguese authors born in the first half of the 18th century would then be explained by the loss of this process. It is difficult to answer the question of whether the loss of V-to-C in BP was triggered by the loss of V-to-C in EP or whether it was independently caused by its own dynamics. In any case, as shown in many synchronic studies, the tendency of the subject to be pre-verbal is stronger in BP than in EP. Figure 10 shows that, after the sharp decrease observed in the texts written by Portuguese authors born at the beginning of the 18th century, the frequency of VS stabilizes at around 25 %, which is similar to what we see in Brazilian authors in the second period of Table 1.

In the subsequent periods, the frequency of post-verbal subjects continues to drop for Brazilian authors, who only use it in 10 % of sentences in the 19th century. This can arguably be attributed to the weakening of the inflectional morphology of the verb and the loss of the consistent null subject property, since it has long been shown that free subject inversion is correlated with consistent null subjects. This difference between the two languages with respect to subject inversion mainly affects the position of subjects of unaccusative verbs, that is, existential, change-of-state, and change-of-place verbs. Santos/Soares da Silva (2012, 134), studying the same corpus of plays as Duarte (2012), show that with definite subjects, VS decreases with such verbs from 49 % in the first period (1838–1845) to 29 % in the last (1990–1992). We will come back to this fact below.

In Gravina's (2014) comparative corpus of newspapers, we observe a similar evolution in Brazilian newspapers, in which VS with unaccusative verbs decreases from 59 % and 60 %, respectively, in the two first periods (1845–1848 and 1890–1898), to 44 % in the third (1945–1948). In Portuguese newspapers, as expected, the rate of VS is constant over the three periods at 56 %, 56 %, and 60 %.

In addition to the fact that BP has weak agreement and became a partial null subject language, which favors SV over VS, the strong tendency for subjects to be pre-verbal may have another cause. In effect, as mentioned above, Santos/Soares da Silva (2012, 134) evidence a difference between definite and indefinite subjects. Only with the former does the rate of inversion change over time. With the latter, it is kept constant. This could be correlated with the topic-oriented property of BP that causes topic phrases to move to the initial position in the sentence and overtly agree with it, although they are not its thematic subject. This very peculiar phenomenon was put in evidence by the pioneering work of Pontes (1987, [1]1981). She exemplifies it with sentences like (35) and (36), in which the genitive phrase *a Sarinha* and the locative

16 Or, alternatively, in a cartographic approach, to one of the heads of the CP layer. Antonelli (2011) argues that this head is Fin.

phrase *estas casas* appear in pre-verbal position, triggering morphological agreement on the verb:

(35) A Sarinha nasceu dois dentes
the Sarinha was-born.3SG two teeth
'Sarinha has got two teeth.'

(36) Estas casas batem muito sol
these houses strike.3PL much sun
'There is too much sun on this house'

This construction, called subject-topic in the literature, has only recently been observed in written texts. But a similar phenomenon appears at the beginning of the 20th century in the corpus of plays studied by Duarte and colleagues. Sentences like (37 a) appear as soon as 1918 (cf. Henriques 2012, 112):

(37) a. O Senhor$_i$ parece que [e]$_i$ anda com más intenções acerca dela.
You seem.3SG that go.3SG with bad intentions about her
'you seem to have bad intentions about her'
b. [e]$_i$ parece que o Senhor$_i$ anda com más intenções acerca dela.
seem that You.3SG go with bad intentions about her
'it seems that you have bad intentions about her'

Two important observations are at stake concerning (37 a). First, it contrasts with (38 b), in which the noun phrase *O Senhor* is in the subject position of the embedded clause. In the first period (1838–1847), Henriques only finds the (b) sentences. Their frequency then gradually decreases in favor of the (a) sentences, falling to 0 % in the fifth period (1949–1955). Second, because of the fact that both verbs are in the singular, the nature of [e] in the subject position of the verb *anda* in (38 a) is ambiguous. Since BP is still a null-subject language, it could be a pronominal null subject. In this case, the NP *O Senhor* would be in topic position, and the subject of *parece* would be an expletive null subject. But [e] could also be the trace left by the movement of *O Senhor* to the matrix subject position. This process, called hyper-raising in the literature, is agrammatical in other languages, like English, but has been frequently found in more recent records of BP, as in (38)—a play written in 1992. Here, the agreement in both matrix and subordinate verbs evidences that *vocês* acts as the subject of both the matrix verb *parecem* and the subordinate verb *pensam*:

(38) Vocês$_i$ parecem que [e]$_i$ não pensam na vida
You.2PL seem.3PL that (you) NEG think about life
'You don't seem to care about life'

The fact that Henriques (2012, 112) finds a sentence like (39) shows that this was already possible in 1918:

(39) Pelo jeito, [e]$_i$ parecem que [e]$_i$ já levam a sua conta
apparently seem.3PL that already take.3PL your account
'It looks like they are already taking your account away'

We can therefore conclude that the topic-oriented grammar characteristic of BP manifests itself in texts as early as the beginning of the 20th century. In the next section, we consider changes affecting objects.

4.2 Objects

As discussed in Section 2.1 (cf. Figure 1), in the BP pronominal paradigm, the clitic pronouns *te* and *lhe* 'you.ACC/DAT' can both refer to the address form *você* 'you'. The striking fact is that, in some BP dialects, they can do so in both direct object and indirect object functions. With respect to the 3rd person form *lhe*, this is a novelty since in Classical Portuguese and in EP, as in many other Romance languages, although the accusative and dative first- and second-person clitic forms are identical, the dative and accusative forms of the third person are distinct. This change is correlated with two other innovations found in BP. First, as already mentioned in Section 2.3, the accusative third person clitic *o*/*a* has disappeared from colloquial speech, and in its place we find either null objects or the strong pronoun *ele*/*ela*. Second, *lhe* ceased to be used in reference to the third person. Instead, we find *ele* preceded by a preposition: *a* 'to' or *para* 'for':

(40) **EP:** João deu-**lhe** o livro
João gave.DAT3SG the book
BP: João deu o livro **a/para ela**
João gave the book to/for her
'John gave her the book'

Figure 11 from Berlinck (2001) shows the evolution of these alternatives in the diachrony of BP:

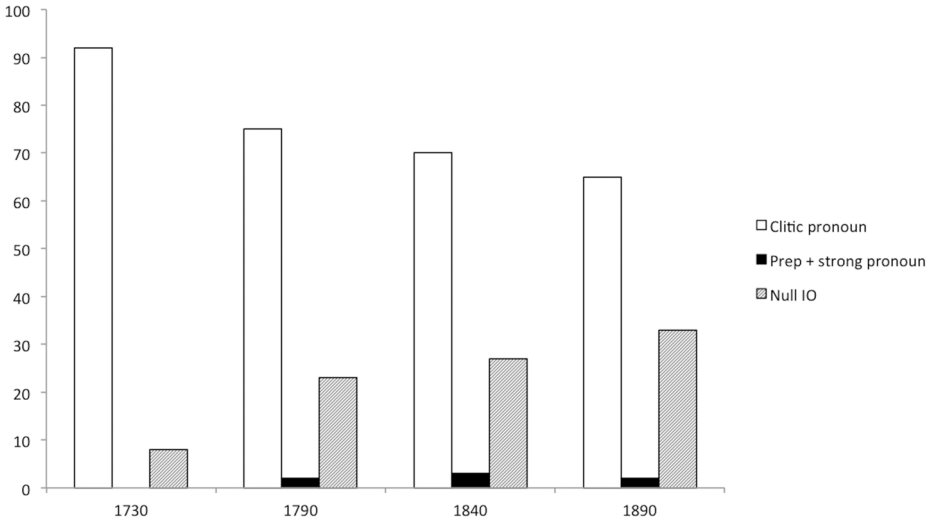


Figure 11: The diachronic development of the expression of Indirect Object (IO) in plays written in BP (Torres/Salles 2010, 189).

Concomitantly, it was observed that, over time, *para* tended to be used instead of *a* to introduce indirect objects. Torres Moraes/Lima Salles (2010) argue that the loss of *lhe* as a third-person dative pronoun and the substitution of *para* for *a* in indirect objects are two reflexes of the same parametric change: the loss of the applicative node, the special functional head that licenses dative arguments, with *a* as the dative marker. As a consequence, indirect objects are expressed by a prepositional phrase whose head is *para*. However, the authors point out that there are dialects of BP in which the applicative node is still active. In such variants, a double object construction is found (cf. Scher 1996; Lucchesi/Baxter/Ribeiro 2009):

(41) ele vendia compade Jacó porco gordo
he sold Father Jacob pig fat
'he sold a fat pig to Father Jacob'
(Lucchesi/Baxter/Ribeiro 2009, 442)

They argue that, in this case, what was lost was the casual distinction between direct and indirect objects. The dative is no longer associated with the applicative node. In historical records, we find a few cases of this construction in the *Proceedings*:[17]

17 Cf. Barros/Figueiredo/Cavalcante (2019).

(42) Senhor Prizidente levou o cuisimento da Sembreia um riquirimento do Senhor Manoel Leonardo
'Mister President brought the knowledge of the Assembly a requirement of Mister Manoel Leonardo'

This is likely not to be accidental. The *Proceedings* were written by ex-slaves, and the double object construction is one of the innovations found in dialectal BP that can be reasonably attributed to the Bantu influence on Portuguese. It is found both in Brazilian Afro-descendent communities and in African Portuguese (Gonçalves 2010). Moreover, Bantu languages have no casual distinction between accusative and dative.

4.3 Relative clauses

Relative clauses, or rather, the differences in the relativization strategies between EP and BP are another aspect recurrently mentioned that syntactically distinguishes the two varieties. Diachronic and synchronic studies argue that the two non-standard relativization constructions, prepositional phrase chopping (PP-chopping) (43) and the resumptive pronoun strategy (44), are becoming productive in BP, the former more frequent than the latter. These constructions coexist with the pied-piping relatives (standard relatives), illustrated in (45):

(43) Nonstandard – PP-chopping
O caminhão **que** eu trabalhava quebrou.
the truck which-REL I work.IMPERF.1SG break.PST.3SG
'The truck which I worked broke down.'

(44) Nonstandard – Resumptive
O caminhão **que** eu trabalhava **com ele** quebrou.
The truck which-REL I work.IMPERF.1SG with it break.PST.3SG
'The truck which I worked with it broke down.'

(45) Standard with pied-piping
O caminhão **com que/o qual** eu trabalhava quebrou.
The truck with which-REL I work.IMPERF.1SG break.PST.3SG
'The truck with which I worked broke down.'

These three types of relativization strategies only occur when the relativized term is a prepositional phrase. In (44), the relative clause starts with the introductory element *que*, while the relativized position is filled by a resumptive pronoun (*ele* 'he' in this case), which refers to the preceding element *o caminhão* 'the truck'. In (43), the posi-

tion filled by the relativized noun phrase in the relative clause is left empty (no resumptive pronoun), and the preposition is also missing, as opposed to the pied-piped strategy (45), in which only the resumptive pronoun is absent, while the preposition is pied-piped.

The emergence of PP-chopping, along with low productivity of the pied-piping strategy in BP, has been a widely explored field from different theoretical perspectives. Tarallo (1996) analyzes this phenomenon as an embedded change, linking it to the rules of pronominal deletion, which first affected the upper syntactic slots (subject and object) and then the lower syntactic slots, such as prepositional phrases, affecting the relative clauses.

In a more recent analysis, Bispo (2014) studied relativization strategies in personal letters and newspaper texts (readers' and editors' letters published in newspapers) which are part of the sample for the project *Para a História do Português Brasileiro* (PHPB —The History of Brazilian Portuguese Project). The table displaying the author's results confirms what had been observed by Tarallo. In manuscript documentation from the 19th century, Bispo (2014) shows a relative increase in PP-chopping between the first and second half of the 19th century: from 4.6 % in the first half (1801–1850) to 8 % in the second half (1851–1900). In the first half of the 20th century, the percentages reach levels above 30 %. No resumptive pronoun data has been found in this sample. Table 2 shows the results of the study of personal letters (19th and 20th century):

Table 2: Distribution of relativization strategies in private letters in the PHPB Project—19th and 20th century— (adapted from Bispo 2014, 226).

Period	1801–1850		1851–1900		1901–1950	
	N	%	N	%	N	%
Standard	42	95.4	230	92.0	148	64.6
PP-chopping	2	4.6	20	8.0	81	35.4
Resumptive pronoun	–		–		–	
TOTAL	44	100.0	250	100.0	229	100.0

In the sample of newspaper texts, however, the percentage of PP-chopping did not reach 10 % in Bispo's study (2014, 227). The percentage of PP-chopping remains stable in the 20th century at 5.9 % in the first half (1801–1850) and 4.5 % in the second half (1851–1900). At the beginning of the 20th century (1901–1950), the author only found 9.5 % PP-chopping. In this sample of newspaper material, as was the case with the manuscript, the author did not locate the resumptive pronoun strategy either.

These results indicate that factors related to the nature of the text (level of formality, communicative purposes, the author's literacy level, etc.) considerably influence

the frequency of the PP-chopping strategy. Private letters maintain traces of communicative proximity between the interlocutors (Koch/Oesterreicher 2007), which surface in the private context; they display strong emotionality, spontaneity, and free topic development. It is worth pointing out that in a study based on letters written in Portugal between the 16th and the 18th centuries, Silva (2011) found PP-chopping frequencies above 10 % on average in her data sample.

Studies based on speech samples show differences in productivity regarding the use of PP-chopping. While in BP, frequencies vary between 70 % and 80 % (see Silva/Lopes 2007; Lessa-de-Oliveira 2006; Barros 2000), the percentage of PP-chopping is around 30 % in EP (34 % in Alexandre 2000; 28 % in Arim/Ramiro/Freitas 2003).

The presence of PP-chopping in the remaining EP documentation could invalidate the hypothesis that PP-chopping is innovative in BP. Nevertheless, the gradual increase of this construction observed in texts written by Brazilians in the 19th century, reaching almost 40 % in the first half of the 20th century (cf. Bispo 2014, 226–227) along with the low rate in EP (10 %) observed by Silva (2011), indicate that the two languages have taken different courses. The stressed quantitative differences suggest that the implementation and diffusion of PP-chopping varies in both territories.

4.4 Interrogative clauses

Interrogative clauses constitute an interesting case in the history of BP because we observe a triple change affecting them over time, as exemplified in (47). First, subjects cease to be obligatorily post-verbal (b). Second, the frequency of the insertion of the copula-complementizer between the WH-phrase and the verb sequence *é que* 'is that' (c) gradually increases and loses the copula, becoming simply *que* (d). Finally, the fronting of the WH-element becomes optional (e).

(45) a. **O que** disse o João?
b. **O que** o João disse?
c. **O que** é que o João disse?
d. **O que** que o João disse?
e. O João disse **o quê**?
'What did John say'

Lopes Rossi (1993) studied the syntax of interrogatives in plays written between 1838 and 1962 as well as soap operas from the second half of the 20th century. She observed that at the beginning, 97 % of the expressed subjects were post-verbal. In the second half of the 19th century, the frequency of post-verbal subjects was still high, but the frequency of the *é que* construction began to increase, giving more space to pre-verbal subjects. At the same time, WH-phrases began to appear in situ. In the 20th century, V-SN dropped to 20 % in the first half and 10 % in the second half of the century.

In this last period, it is possible to compare plays and soap operas. In both cases, preverbal subjects are largely predominant. However, some interesting differences appear. First, *que*, the reduced version of *é que*, is absent from plays but corresponds to half of the cases of (*é) que* in soap operas. Second, WH-in situ is much more frequent in soap operas (31 % of the occurrences, against 8 % in plays). Finally, null objects are more frequent in plays (12.5 % vs. 2 % in soap operas). In sum, we see that the 4 innovations exemplified in (45 b–e) are present in both kinds of texts, but, quantitatively, soap operas are closer to present-day colloquial speech. This could be a matter of genre, though more recent work and data suggest that it is also a question of time.

In Duarte's (2012) corpus of plays previously mentioned in this chapter, Pinheiro/Marins (2012) observe that there is a strong increase in the frequency of *é que* interrogative clauses between the penultimate (1975) and final (1992) periods (23 % to 58 %).[18] As for the *que* constructions, they are absent from the corpus until 1975, appearing then with a frequency of 16 %. Since the last play studied by Lopes Rossi is from 1962, the complete absence of *que* interrogatives in her corpus is likely due to this construction being too recent in speech to be reproduced in a literary work.[19]

5 Concluding remarks

In this chapter, we reviewed the main morpho-syntactic changes observed in the history of Brazilian Portuguese. Although written documents are not the best means for analysing linguistic innovation, we showed that the emergence and evolution of the syntactic peculiarities of BP can be detected in various texts such as newspapers, plays, advertisements, letters, and meeting proceedings during the 19th and 20th centuries. Most of the peculiarities clearly appear in the second half of the 19th century, reinforcing the claim put forth by Tarallo (1996) that BP turned different at the turn of the 19th century.

These linguistic changes are generally a long process and, in most cases, have not yet reached completion. Anaphoric objects, for instance (cf. Figure 4), are consistently less often realized as clitics and more expressed as null objects over time. In our records, this tendency first appears between 1850 and 1900 and continues to be observed in the first half of the 20th century. However, at the end of the period considered in this essay (1945), clitic pronouns still were the preferred choice in written text. We saw that the tendency is the opposite for subjects in the sense that null subjects tend to be substituted by pronominal subjects (cf. Figure 6). In this case, however, we can follow the dynamics of the change over a longer span of time. Interestingly, such dynamics vary according to the person of the pronoun. For the third person, we observe that the use of

18 The frequencies cannot be fully compared because Pinheiro/Marins do not consider WH-in situ.
19 It must be noted, however, that the frequency of null subjects is closer to Lopes Rossi's last play.

lexical pronouns increases between the first and second half of the 19th century and then remains stable for nearly 50 years, increasing again between 1950 and the end of the 20th century. The second-person pronouns, in turn, increase more slowly at the beginning then much faster between 1920 and 1950, arriving then at a rate that is stable until the end of the century. The first-person pronouns evolved in yet another fashion.

Beyond distinctions in frequency rates, new grammaticalized pronouns, such as *você* (2SG) and *a gente* (1PL), are also involved in relevant changes in the BP syntactic structure: weakening of the verbal morphology and partial loss of the null subject parameter. It therefore appears that the particular dynamics of each change in the texts are due to multiple factors that have to be more thoroughly investigated. A few changes appear earlier or later than the second half of the 19th century. The inflection of the use of the periphrastic future, for instance, appears in the 18th century (cf. Figure 8), which is also the case for the indirect pronominal objects realized as Preposition + Strong pronoun (cf. Figure 11). Conversely, the use of existential *ter* only manifests itself from the 1940s on. Finally, some of the constructions considered in this chapter as being the locus of syntactic changes are too poorly, or too recently, represented in texts to be analyzed from the point of view of their quantitative evolution. This raises the issue of whether such changes were not visible because of the pressure of language norms or because they are genuinely recent innovations. This question can only be answered by gathering and analysing more data.

6 References

Alexandre, Nélia Maria Pedro (2000), *A Estratégia Resumptiva em Relativas Restritivas do Português Europeu*, Master's thesis, Lisboa, Universidade de Lisboa.

Antonelli, André (2011), *Sintaxe de posição do verbo e mudança gramatical na história do português europeu,* Doctoral thesis, Campinas, Universidade Estadual de Campinas.

Arim, Eva/Ramiro, Maria Celeste/Freitas, Tiago (2003), *Estratégias de relativização nos meios de comunicação social portugueses*, Actas do XIX encontro Nacional da Associação Portuguesa de Linguística, Lisboa, APL, 279–288, https://apl.pt/wp-content/uploads/2017/12/2003-21.pdf, last accessed 16.02.2022.

Avelar, Juanito (2009), *On the Emergence of TER as an Existential Verb in Brazilian Portuguese*, Historical Syntax and Linguistic Theory 1, 158–175.

Avelar, Juanito/Callou, Dinah (2007), *Sobre a emergência do verbo possessivo em contextos existenciais no português brasileiro*, in: Ataliba Castilho et al. (edd.), *Descrição, história e aquisição do português brasileiro*, São Paulo, Pontes, 375–402.

Avelar, Juanito/Cyrino, Sonia (2014), *Construções com SE e Constituintes Locativos no Português Brasileiro: Indagações e Hipóteses nos Eixos Sincrônico e Diacrônico*, Paper presented at GSCP International Conference, Stockholm University.

Avelar, Juanito/Galves, Charlotte (2014), *O papel das línguas africanas na emergência do português brasileiro,* Linguística 30/2, 241–288.

Barros, Alvanira L. de (2000), *O uso da relativa cortadora na fala pessoense*, Master's thesis, João Pessoa, Universidade Federal da Paraíba.

Barros, Isis/Figueiredo Cristina/Cavalcante Rerisson (2019) *A expressão do dativo*, in: Charlotte Galves/Tânia Lobo (edd.) *O português escrito por afro-brasileiros no século XIX. As atas da Sociedade Protetora dos Desvalidos*, Salvador, Edufba, 175–216.

Berlinck, Rosane de Andrade (2000), *Brazilian Portuguese VS Order: a diachronic analysis*, in: Mary A. Kato/Esmeralda V. Negrão (edd.), *Brazilian Portuguese and the Null Subject Parameter*, Frankfurt/Madrid, Vervuert/Iberoamericana, 75–195.

Berlinck, Rosane de Andrade (2001), *Dativo ou locativo? Sobre sentidos e formas do "dativo" no português,* Letras 56, 159–175.

Biberauer, Theresa, et al. (edd.) (2010), *Parametric Variation: Null subjects in Minimalist Theory*, Cambridge, Cambridge University Press.

Bispo, Edvaldo (2014), *Orações Relativas em Perspectiva Histórica: Interface Uso e Cognição*, Veredas 18, 222–235.

Brocardo, Maria Teresa (2014), *Tópicos de História da Língua Portuguesa*, Lisboa, Colibri.

Callou, Dinah/Barbosa, Afranio G. (edd.) (2011), *A Norma Brasileira Em Construção: Cartas A Rui Barbosa (1866 a 1899)*, 2 vol., Rio de Janeiro, Fundação Casa de Rui Barbosa.

Carneiro, Zenaide (2005), *Cartas da Bahia (1809–1907): um estudo filológico-linguístico*, Doctoral thesis, Campinas, Universidade Estadual de Campinas.

Carneiro, Zenaide/Galves, Charlotte (2010), *Variação e Gramática: colocação de clíticos na história do português brasileiro*, Revista de Estudos da Linguagem 18/1, 7–38.

Carpani, Marina (2010), *Sujeito nulo e sujeito pronominal lexical no Corpus Histórico do Português Tycho Brahe*, IC Report, Campinas, Universidade Estadual de Campinas.

Castilho, Célia (2013), *Fundamentos sintáticos do português brasileiro*, São Paulo, Contexto.

Chomsky, Noam (1995) *The Minimalist Program*, Cambridge, Massachusetts, The MIT Press.

Cintra, Luís Filipe Lindley (1972), *Sobre "Formas de Tratamento" na Língua Portuguesa*, Lisboa, Livros Horizonte.

Clackson, James/Horrocks, Geoffrey (2011), *The Blackwell History of the Latin Language*, Oxford, Wiley-Blackwell.

Coelho, Izete L./Gorski, Edair (2011), *A Variação no Uso dos Pronomes "Tu" e "Você" em Santa Catarina*, in: Célia Lopes/Letícia Rebollo (edd.), *Formas de tratamento em Português e Espanhol: Variação, Mudança e Funções Conversacionais*, Niterói, Editora da UFF, 263–287.

Correa, Vilma (1991), *O objeto direto nulo no português do Brasil*, Master's thesis, Campinas, Universidade Estadual de Campinas.

Cyrino, Sonia (1997), *O objeto nulo no português brasileiro: um estudo sintático-diacrônico*, Londrina, Editora da UEL.

Duarte, M. Eugênia L. (1986), *Variação e sintaxe: clítico acusativo, pronome lexical e categoria vazia no português do Brasil*, Master's thesis, São Paulo, PUC.

Duarte, M. Eugênia L. (1995), *A Perda do Princípio "Evite Pronome" no Português Brasileiro,* Doctoral thesis, Campinas, Universidade Estadual de Campinas.

Duarte, M. Eugênia L. (2012), *Os sujeitos de 3a pessoa: revisitando Duarte 1993*, in: M. Eugênia L. Duarte (ed.), *O sujeito em peças de teatro (1833–1892). Estudos diacrônicos*, São Paulo, Parábola, 21–44.

Fraga, I. P. de/Gomes, Valeria Severino (2014), *Cartas pessoais—Recife, Pernambuco*, Recife, Projeto PHPB/PE.

Galves, Charlotte (2001, [1]1987), *A sintaxe do português brasileiro*, Estudos de Linguística 13, 31–49. (Reprinted in: *Ensaios sobre as gramáticas do português*, Campinas, Editora da Unicamp).

Galves, Charlotte (1993), *O Enfraquecimento da Concordância no Português Brasileiro*, in: Ian Roberts/Mary Kato (edd.), *Português brasileiro, uma viagem diacrônica*, Campinas, Editora da Unicamp, 387–403.

Galves, Charlotte (2018), *Ainda sobre os pronomes do português brasileiro:sintaxe, morfologia e variação*, in: Alessandro Boechat de Medeiros/Andrew Nevins (edd.) *O apelo às árvores. Estudos em homenagem a Miriam Lemle*, Campinas, Pontes, 79–100.

Galves, Charlotte (2019), *Competition, stability and change in the emergence of Brazilian Portuguese*, in: Anne Breitbarth/Miriam Bouzouita/Lieven Danckaert/Melissa Farasy (edd.), *The Determinants of Diachronic Stability*, Amsterdam, Benjamins, 192–214.

Galves, Charlotte (in preparation), *European Portuguese and Brazilian Portuguese: how quickly may languages diverge?*, in: Paola Crisma/Giuseppe Longobardi (edd.), *The Oxford Handbook of Diachronic and Historical Linguistics*, Oxford, Oxford University Press.

Galves, Charlotte/Britto, Helena/Paixão de Sousa, Maria C. (2005), *The change in clitic placement from Classical to Modern European Portuguese: results from the Tycho Brahe Corpus*, Journal of Portuguese Linguistics 4/1, 39–68.

Galves, Charlotte/Gibrail, Alba (2018), *Subject inversion in transitive sentences from Classical to Modern European Portuguese: a corpus-based study*, in: Adriana Cardoso/Ana M. Martins (edd.) *Word Order Change*, Oxford, Oxford University Press,163–178.

Galves, Charlotte/Lobo, Tânia (2019), Os clíticos pronominais, in: Charlotte Galves/Tânia Lobo (edd.) *O português escrito por afro-brasileiros no século XIX. As atas da Sociedade Protetora dos Desvalidos*, Salvador, Edufba, 121–174.

Galves, Charlotte/Paixão de Sousa, Maria C. (2017), *The change in the position of the verb in the history of Portuguese: Subject realization, clitic placement and prosody*, Language 93/3, 152–180.

Galves, Charlotte/Ribeiro, Ilza/Torres Moraes, Maria A. (2005), *Syntax and morphology in the placement of clitics in European and Brazilian Portuguese*, Journal of Portuguese Linguistics 4/2, 143–177.

Galves, Charlotte/de Andrade, Aroldo/Faria, Pablo (2017), *The Tycho Brahe Corpus of Historical Portuguese*, http://www.tycho.iel.unicamp.br/~tycho/corpus/en/index.html, last access 16.02. 2022.

Gomes, Valéria Severino/Lopes, Célia Regina (2014), *Variação entre Formas dos Paradigmas de "Tu"—"Você" em Cartas Pernambucanas dos Séculos XIX e XX*, Revista do GELNE 16, 1–15.

Gonçalves, Perpétua (2010), A gênese do português de Moçambique, Lisboa, INCM.

Gravina, Aline (2008), *A natureza do sujeito nulo na diacronia do PB: estudo de um corpus mineiro (1854–1950),* Master's thesis, Campinas, Universidade Estadual de Campinas.

Gravina, Aline (2014), *Sujeito nulo e ordem VS no português brasileiro: um estudo diacrônico-comparativo baseado em corpus*, Doctoral thesis, Campinas, Universidade Estadual de Campinas.

Guy, Gregory (1981), *Linguistic variation in Brazilian Portuguese: aspects of phonology, syntax and language history*, Doctoral thesis, Philadelphia (PA), University of Pennsylvania.

Henriques, Fernando Pimentel (2012), *Construções com verbos de alçamento que selecionam um complemento oracional*, in: M. Eugênia L. Duarte (ed.), *O sujeito em peças de teatro (1833–1892). Estudos diacrônicos*, São Paulo, Parábola, 101–120.

Kato, Mary A./Negrão, Esmeralda V. (edd.) (2000), *Brazilian Portuguese and the Null Subject Parameter,* Madrid/Frankfurt, Iberoamericana/Vervuert.

Koch, Peter/Oesterreicher, Wulf (2007), *Lengua hablada en la Romania: español, francés, italiano*, Madrid, Gredos.

Lessa-De-Oliveira, Adriana Stella C. (2006), *Dados de Mudança no Sistema de Relativização em Português Brasileiro,* Estudos Lingüísticos 35, 224–233.

Lopes, Célia Regina dos Santos (2019), *A formação dos sistemas de tratamento em português: mudança e avaliação*, LaborHistórico 5 (Especial 1), 257–294.

Lopes, Célia Regina dos Santos (2003), *A inserção de "A gente" no Quadro Pronominal do Português,* Frankfurt am Main/Madrid, Vervuert/Iberoamericana.

Lopes, Célia Regina dos Santos (ed.) (2005), *A norma Brasileira em Construção: Fatos Lingüísticos em Cartas Pessoais do Século XIX*, Rio de Janeiro, UFRJ/FAPERJ.

Lopes, Célia Regina dos Santos/Marcotulio, Leonardo Lennertz (2019). *On address pronouns in the history of Brazilian Portuguese*, in: Bettina Kluge; María Irene Moyna, *It's not all about you: New perspectives on address research*, Amsterdam, Benjamins, 140–159.

Lopes, Célia Regina dos Santos/Souza, Camila/Marcotulio, Leonardo/Oliveira, Thiago/Lucena, Rachel (2018), *A reorganização no sistema pronominal de 2a. pessoa na história do português brasileiro: outras relações gramaticais*, in: Célia Regina dos Santos Lopes (ed.), *História do português brasileiro*, vol. 4: *Mudança sintática das classes de palavra: perspectiva funcionalista*, São Paulo, Editora Contexto, 106–189.

Lopes, Célia Regina dos Santos/Cavalcante, Silvia Regina (2011), *A Cronologia do Voceamento no Português Brasileiro: expansão de "Você"-sujeito e Retenção do Clítico "-te",* Revista Lingüistica 25, 30–65.

Lopes, Célia Regina dos Santos/Duarte, Maria Eugenia Lamoglia (2003), *De "Vossa Mercê" A "Você": Análise Da Pronominalização De Nominais Em Peças Brasileiras E Portuguesas Setecentistas E Oitocentista*s, in: Silvia F. Brandão/Maria Antonia Mota (edd.), *Análise Contrastiva de Variedades do Português*, vol. 1, *Primeiros Estudos*, Rio de Janeiro, In-Fólio, 61–76.

Lopes, Célia Regina dos Santos/Rumeu, Marcia Cristina de Brito/Carneiro, Zenaide (2013), *A Configuração Diatópico-Diacrônica do Sistema de Tratamento do Português Brasileiro*, Revista do GELNE 15, 187–212.

Lopes, Célia Regina dos Santos/Souza, Camila Duarte/Oliveira, Thiago Laurentino (2013), *A frequência e o Delineamento da Gramática: a Afixação do Clítico Te no Português Brasileiro*, Veredas (UFJF. Online) 17, 376–397, https://periodicos.ufjf.br/index.php/veredas/article/view/25010, last accessed 16.02.2022.

Lopes Rossi, Maria A. (1993), *Estudos diacrônicos sobre as interrogativas do português do Brasil*, in: Ian Roberts/Mary Kato (edd.), *Português brasileiro, uma viagem diacrônica,* Campinas, Editora da Unicamp, 307–342.

Lucchesi, Dante/Baxter, Alan/Ribeiro, Ilza (edd.) (2009), *O Português Afro-Brasileiro*, Salvador, Edufba.

Macedo Costa, Tatiane (2012), *Um estudo diacrônico das variadas realizações do objeto direto anafórico na imprensa baiana dos séculos 19 e 20*, Master's thesis, Campinas, Universidade Estadual de Campinas.

Marcotulio, Leonardo L. (2010), *Língua e História: O 2º Marquês do Lavradio e as Estratégias Linguísticas da Escrita no Brasil Colonial*, Rio de Janeiro, Ítaca.

Marins, Juliana (2013), *As Repercussões da Remarcação do Parâmetro do Sujeito Nulo: um Estudo Diacrônico das Sentenças Existenciais com Ter e Haver no PB e no PE*, Doctoral thesis, Rio de Janeiro, UFRJ.

Martins, Marco Antonio/Moura, Kássia Kamilla (2013), *Investigando a Influência do Contexto Morfossintático na Implementação de "Você" em Cartas Particulares do Rio Grande do Norte no Século 20*, Revista do GELNE 15, 245–266.

Mattos e Silva, Rosa V. (1997), *Observações sobre a Variação no Uso dos Verbos "Ser", "Estar", "Haver", "Ter" no Galego-Português Ducentista*, Estudos lingüísticos e literários 19, 253–285.

Mattos e Silva, Rosa V. (2004), *Ensaios para uma sócio-história do português brasileiro*, São Paulo, Parábola Editorial.

Naro, Anthony/Scherre, Marta (2007), *As origens do português brasileiro*, São Paulo, Parábola Editorial.

Negrão, Esmeralda/Viotti, Evani (2000), *Brazilian Portuguese as a Discourse-oriented language*, in: Mary A. Kato/Esmeralda V. Negrão (edd.), *Brazilian Portuguese and the Null Subject Parameter*, Madrid/Frankfurt, Iberoamericana/Vervuert, 105–125.

Nunes de Souza, Christiane/Coelho, Izete (2013), *O Sistema de Tratamento em Santa Catarina: Uma Análise de Cartas Pessoais dos Séculos XIX e XX*, Revista do GELNE 15, 213–243.

Oliveira, Josane Moreira (2006), *O Futuro da Língua Portuguesa Ontem e Hoje: Variação e Mudança*, Doctoral thesis, Rio de Janeiro, UFRJ.

Oliveira, Klebson (2006), *Negros e escrita no Brasil do século XIX: sócio-história, edição filológica de documentos e estudo linguístico*, Doctoral thesis, Salvador, Universidade Federal da Bahia.

Paixão de Sousa, Maria C. (2004), *Língua barroca: Sintaxe e história do português nos 1600,* Doctoral thesis, Campinas, Universidade Estadual de Campinas.

Pagotto, Emílio (1992) *A posição dos clíticos em português: um estudo diacrônico*, Master's thesis, Campinas, Universidade Estadual de Campinas.

Pinheiro, Diogo/Marins, Juliana (2012), *A trajetória das Interrogativas QU- clivadas e não clivadas no português brasileiro*, in: M. Eugênia L. Duarte (ed.), *O sujeito em peças de teatro (1833–1892). Estudos diacrônicos*, São Paulo, Parábola, 161–180.

Pontes, Eunice (1987, [1]1981), *Da importância do tópico em português*, in: *Anais do 5o Encontro de Linguistica*, Rio de Janeiro, PUC. (Reprinted in: *O tópico no português do Brasil*, Campinas, Pontes).

Ribeiro, Ilza (1998), *A mudança sintática do português brasileiro é mudança em relação a que gramática?*, in: Ataliba T. Castilho (ed.), *Para a história do português brasileiro: primeiras idéias*, São Paulo, Humanitas, 101–120.

Rumeu, Márcia (2013), *Língua e Sociedade: a História do Pronome "Você" no Português Brasileiro*, Rio de Janeiro, Ítaca.

Rumeu, Marcia Cristina (2014), *A Difusão do "Você" pelos Contextos Sintáticos de Complementação e de Adjunção*, Revista Portuguesa de Humanidades 18/I, 91–114.

Santos, Danielle de Rezende/Soares da Silva, Humberto (2012), *Ordem V-DP/DP-V com verbos inacusativos*, in: M. Eugênia L. Duarte (ed.), *O sujeito em peças de teatro (1833–1892). Estudos diacrônicos*, São Paulo, Parabola, 121–142.

Scher, Ana P. (1996), *As construções com dois complementos no inglês e no português do Brasil: um estudo sintático-comparativo*, Master's thesis, Campinas, Universidade Estadual de Campinas.

Scherre, Marta, et al. (2015), *Variação dos Pronomes "Tu" e "Você"*, in: Marco Antonio Martins/ Jussara Abraçado, *Mapeamento Sociolinguístico do Português Brasileiro*, São Paulo, Contexto, 133–172.

Silva, Bianca G. S. G. (2011), *As estratégias de relativização na escrita de portugueses dos séculos XVI, XVII e XVIII*, Doctoral thesis, Rio de Janeiro, UFRJ.

Silva, Bianca Graziela/Lopes, Célia (2007), *O Papel da Freqüência na Gramaticalização do Que: Análise das Estratégias de Relativização no Português do Brasil*, Veredas 11, 80–100.

Souza, Camila Duarte/Oliveira, Thiago Laurentino (2013), *A Representação da 2ª Pessoa nas Posições de Complemento: o Papel da Categoria Social*, Working Papers em Linguística (Online) 14, 100–120. DOI: https://doi.org/10.5007/1984-8420.2013v14n2p100.

Tarallo, Fernando (1983), *Relativization strategies in Brazilian Portuguese*, Doctoral thesis, Philadelphia (PA), University of Pennsylvania.

Tarallo, Fernando (1993), *Sobre a alegada origem crioula do português brasileiro: mudanças sintáticas aleatórias*, in: Ian Roberts/Mary Kato (edd.), *Português brasileiro, uma viagem diacrônica, Campinas*, Editora da Unicamp, 35–68.

Tarallo, Fernando (1996), *Turning Different at the Turn of the Century: 19th Century Brazilian Portuguese*, in: Gregory R. Guy et al. (edd.), *Towards a Social Science of Language. Papers in honor of William Labov*, vol 1: *Variation and change in language and society*, Amsterdam/ Philadelphia, Benjamins, 199–220.

Torres Moraes, Maria A. (1993), *Aspectos diacrônicos do movimento do verbo, estrutura da frase e caso nominativo no português do Brasil*, in: Ian Roberts/Mary Kato (edd.), *Português brasileiro, uma viagem diacrônica*, Campinas, Editora da Unicamp, 263–306.

Torres Moraes, Maria A. (2002), *Rastreando aspectos gramaticais e socio-históricos do português brasileiro em anúncios de jornais do século XIX*, in: Tânia Alkmim (ed.), *Para a história do português brasileiro*, vol. 3: *Novos Estudos*, São Paulo, Humanitas, 69–126.

Torres Moraes, Maria A./Lima Salles, Heloisa (2010), *Parametric change in the grammatical encoding of indirect objects in Brazilian Portuguese*, Probus 22, 181–209.

Américo Venâncio Lopes Machado Filho

7 The history of the lexicon

Abstract: This chapter outlines the history of the Brazilian Portuguese lexicon from the first decades of colonisation to the end of the 20th century. It begins by describing some of the most useful sources and instruments for historical-lexical research (dictionaries, corpora, current projects) as well as a number of pertinent issues relating to the general architecture of the lexicon (common lexicon, variety-specific lexicon, specialized lexicon). The historical evolution of the Brazilian Portuguese lexicon is seen here from a diachronic perspective, with special attention to lexical innovations, in addition to borrowings introduced by the colonisers (including remarks here on their diastratic and diatopic characterisation) and by Indigenous and Afro-Brazilian communities. Lexical contributions of other languages, including indirect loans, are also noted. The question of lexical unity and diversity in the Lusophone world is also discussed.

Keywords: historical lexical-research in Brazil, diachronical-lexical architecture, historical evolution of Brazilian lexicon, lexical borrowings, Brazilian lexicon and the Lusophone world

1 Foreword

The great Brazilian poet and writer, Carlos Drummond de Andrade (1985, 20), claims through the voice of a character in one of his short stories, that 'you shouldn't plagiarise eternity'.[1] This poetic aphorism was based on the idea of the creation of a "starting point" for the verb *prorrogar* (which means 'to extend'), as if from then on everything could be "extended", and there would be neither change nor history.

However, the "false eternity" invoked by this ending of limits is in fact defeated by the author himself early in his work, when he makes "the signs of the provisory and the contingent" become once more irrefutable. The lesson of the poet lends itself to a demonstration of the paradoxical character that the lexicon assumes in the process of the identity construction of a language. If, on the one hand, it retains the secular extension of its most distant etymological sources, which have solidified themselves in their historical basis, on the other, it allows *the new* to intervene due to cultural or linguistic contacts that this same language in use is exposed to, and to neological processes awoken by the expressive necessities of speech communities.

1 "não se deve plagiar a eternidade."

https://doi.org/10.1515/9783110405958-008

The historical architecture of the lexicon adapts itself to a foundation of great complexity, inseparably structured in terms of its social-historical make-up—just as happens on other formative levels of a language. However, this tendency shows itself in the lexicon in a much clearer way than in morphology or syntax, since we might assume that it is in the lexical inventory

> 'that the designations that label the chain of changes in the ways and journeys of humanity are recorded, besides composing the scenery of revelation for both the reality and the cultural facts that permeated their history' (Barcelos da Silva 2000, 142).[2]

This chapter provides an account of the scientific treatment of this complex socio-historic reality-in-the-making in Brazilian Portuguese, first, by introducing the different approaches that have produced significant results in this endeavour, in section 2, followed by a discussion of current methodological standards of research, in section 3. Section 4 then provides a comprehensive history of the Brazilian Portuguese lexicon, and section 5 integrates this history into more recent divergent and convergent developments in the Lusophone world.

2 Historical-lexical research in Brazil today

From an epistemological perspective, it can be said that, after a lengthy period of ostracism, lexical studies have returned to the spotlight of contemporary Linguistics in Brazil. Many studies of a historical and variational perspective have emerged, and recently some of these, through the use of corpora of trusted philological editions or scientifically collected data, have sought out different moments in the shaping of the Brazilian lexicon.

However, we must never lose sight of the fact that reality in its totality will always be beyond our reach. And, as Piel (1989) affirms,

> 'it will never be possible to reconstitute all the stages that [the lexicon] has explored, and pinpoint the contribution from the many generations that have collaborated in the construction of the impressive edifice we see nowadays in the great modern dictionaries' (Piel 1989, [1]1976, 9).[3]

But even with such limitations in the humanities, one must highlight, among many other noteworthy projects, one of great relevance to the historical re-composition of Brazil's lexicon: the *Projeto Dicionário Histórico do Português do Brasil-séculos XVI,*

---

2 "que se gravam—e não raro, pirogravam—as designações que rotulam as mudanças encadeadoras dos caminhos e dos descaminhos da humanidade, além de comporem o cenário da revelação tanto da realidade quanto dos fatos culturais que permearam sua história".

3 "nunca será possível reconstituir todas as fases por ele percorridas e destrinçar a contribuição das muitas gerações que nele colaboraram até se constituir o magno edifício que hoje se nos depara nos grandes dicionários modernos".

XVII e XVIII (DHPB),[4] conceived and started in 2005 by one of the best-known Brazilian lexicographers, Maria Tereza Biderman, from the Universidade Estadual Paulista (UNESP).

Supported by the *Conselho Nacional de Desenvolvimento Científico e Tecnológico* (CNPq),[5] and with considerable funds from the *Programa Instituto do Milênio*,[6] which had at that point chosen only 17 projects from different scientific areas for financial support, the DHPB was concluded in 2014 under the coordination of Clotilde Murakawa, from the same university, working with a team of specialists from different fields, including linguists, lexicographers, philologists (especially in the first phase of *corpora* selection), and IT professionals, among others.

The project encompassed research in a period of three centuries, and registered Brazil's lexicon in different areas, such as flora, fauna, uses and customs, food, diseases, etc, in this way aiming to fill the gap in our knowledge about the Brazilian lexicon prior to the process of Brazil's political independence. The research resulted in 10,470 entries, collected in 19 volumes, a considerable amount of material which, if printed on A4 paper, would yield over 11 thousand pages. It has not yet been published, but negotiations are ongoing towards this goal.

Before this research, previous efforts on the same lines had been made, such as those of Antônio Joaquim Macedo Soares, who in 1888 published a first attempt at describing Portuguese vocabulary in Brazil, the *Diccionario Brazileiro da Língua Portuguesa*. According to Biderman ([2]2001, 71) only the entries as far as the letter C were finished at that time, and Soares' work was completed over half a century later by his son, based on his father's notes, and was published by the *Instituto Nacional do Livro* (INL)[7] (cf. Soares 1954).

Another notable research project related to the lexicon has been developed since 2012 by *Núcleo de apoio à pesquisa em Etimologia e História da Língua Portuguesa*[8] (NEHiLP), from the University of São Paulo (USP). It brings together many specialists in language history, morphology, lexicography and etymology from national and international institutions, and is coordinated by Mário Viaro, from the same university. From 2015 the group has been working on the challenge of building the *Dicionário Etimológico da Língua Portuguesa*[9] (DELPo), Portuguese being one of the modern Romance languages that still lacks such a dictionary:

> 'Compared to the other Romance languages, Portuguese is the one that has by far the most fragmented and incomplete etymological information. Etymological dictionaries have frequently

4 Historical Dictionary of Brazilian Portuguese Project—16th, 17th and 18th centuries.
5 National Council for Scientific and Technological Development.
6 Millennium Institute Program.
7 National Book Institute.
8 Research support nucleus for Etymology and History of the Portuguese Language.
9 Etymological Dictionary of the Portuguese Language.

> used vague or inaccurate data about the oldest of their entries. At best, such works have collected the knowledge of many etymologists who trust their own research. Yet this kind of attitude forestalls a deeper exploration of the ancient stages of the Portuguese language' (Viaro et al. 2015, 15).[10]

The construction of DELPo assumes even greater importance when considered from within the *Projeto Para a História do Português Brasileiro*[11] (PHPB), and also by conducting etymological research as a means of backdating vocabulary units, this in order to establish their *termini a quo and ad quem*, that is, the first and last lexical occurrences in the *corpora* analyzed. All this research and data analysis is supported by different computational tools such as fragmentation and concordance programs, which are indispensable for lexicographical work today.

The *Projeto Léxico Histórico do Português Brasileiro*,[12] based in the Universidade Estadual de Londrina (UEL) and related to the *Projeto para a História do Português Brasileiro* (PHPB), has as its objective the collection of vocabulary present in manuscripts from ancient villages of the captaincies and governorships of Paraná, Rio de Janeiro, Bahia and Paraíba. Developed by a team coordinated by Vanderci Aguilera (UEL), it focuses on official documents from the 17th, 18th and 19th centuries in semi-diplomatic editions, produced and analyzed by each group of researchers for one of those states.

In parallel to this, other sub-projects are being developed, among them the *Léxico Histórico do Português do Paraná*[13] (LHisPar), which will also be integrated into the PHPB. In this project, 730 manuscript folios from Paraná were used, these coming from the ancient villages of Curitiba, Paranaguá, Castro, Morretes and Lapa, among others, and dating from the 18th and 19th centuries. The LHisPar is now concluded and is currently undergoing a review phase, and will be published on the Internet with about 4,200 entries. Most work has now been completed, but the historical lexicon of Paraíba, Bahia and Rio de Janeiro are still being reviewed.

Another lexicon investigation project in Brazil is the *Dicionário Dialetal Brasileiro*[14] (DDB). It is associated with the *Atlas Linguístico do Brasil*[15] (ALiB) project at the Federal University of Bahia. The DDB, through the use of lexicological and lexicogra-

---

10 "Comparado com outras línguas românicas, o português é, de longe, o idioma que tem a informação etimológica mais incompleta e fragmentada. Os dicionários etimológicos têm, com frequência, informação muito vaga e imprecisa sobre as ocorrências mais antigas dos seus verbetes. Na melhor das hipóteses, estas obras reúnem o conhecimento de vários etimólogos e confiam em suas supostas especializações. Essa atitude, todavia, impede um maior aprofundamento do conhecimento das antigas etapas da língua portuguesa".

11 Project Towards the History of Brazilian Portuguese.

12 Historical Brazilian Portuguese Lexicon Project.

13 Portuguese Historical Lexicon from Paraná.

14 Brazilian Dialectal Dictionary.

15 Linguistic Atlas of Brazil Project.

phical methodologies, aims to interpret the answers that informants gave to three questionnaires used by the ALiB team for data collection. These are the *Questionário Fonético-Fonológico* (QFF), a phonetic-phonological questionnaire incorporating 159 questions, the *Questionário Semântico-Lexical* (QSL), a semantic-lexical questionnaire with 202 questions distributed over 14 thematic and conceptual areas, and the *Questionário Morfossintático* (QMS), a morphosyntactic questionnaire with 49 questions.

Considering that this study takes a variational perspective, the structure of DDB will allow the recognition of spatial and social varieties registered by ALiB for each of the variables that the research is concerned with, thus allowing for fast access from a microstructure capable of condensing such information, including the phonetic variations from lexical items, in a practical and economical manner. It will also allow the identification of possible lexical and semantic relations among the registered items (cf. Machado Filho 2010). The construction of dialectical vocabularies from Bahia (VDB) and Centro-Oeste Brasileiro (VDCO) are already at an advanced stage, and their results will be integrated to the DDB.

Apart from at the strictly lexical level, that is, studies that are directed to the representational lexical units of the language, research into terminology has yielded significant results, which are reflected in the expansion of the production of terminological vocabularies, so necessary in professions such as medicine, law, technology, the environment, linguistic studies, and even politics, among others. In this vein we might mention the research of the Projeto Terminológico Cone Sul[16] (TERMISUL), whose work began in 1991 under the leadership of Maria da Graça Krieger, originally from the Federal University of Rio Grande do Sul (UFRGS). This project has proceeded on the lines of theoretical and applied research into terminology, assuming the perspective of specialized languages manifested in texts concerning the terminological area under investigation.

3 Methodological standards and terminological distinctions

The projects mentioned above demonstrate some of the different approaches through which the lexicon has been investigated and recorded, either within the dimension of the common cultural and identity assets of a language—in a historical perspective—, as a characterising element of certain dialects or sociolects of that language—under the sociolinguistic and dialectal approach—, or even from the perspective of specialized languages.

16 Terminological Project of the South Cone.

Considering the latter, Dubois et al. (1994, 440) state:

> On appelle langue de spécialité un sous-système linguistique tel qu'il rassemble les spécificités linguistiques d'un domaine particulier. En fait, la terminologie, à l'origine de ce concept, se satisfait très généralement de relever les notions et les termes considérés comme propres à ce domaine. Sous cet angle, il y a donc abus à parler de langues de spécialité, et vocabulaire spécialisé convient mieux.[17]

Now, the notions of "subsystem" and "specialized language", opposed to "specialized vocabulary", refer to a reflection about the borders that concepts associated with the specialized lexicon can establish with respect to the "common lexicon" or the "lexicon of the language", and the lexicon which can be called "socio-dialectal" or "normative lexicon".

The crucial question that is bound up with such notions lies, in fact, under the aegis of four main aspects that need to be borne in mind from a lexical perspective:

1. The precise definition of the theoretical object, that is, the treatment unit from the lexical-lexicographical point of view which, in cases of 'common lexicon' and 'specialized lexicon', is distributed along the continuum between the 'lexie'[18] and the 'term'. It is possible to say that strongly characterising records of sociodialectical norms could, on the same continuum, be distributed under the label that here we shall refer to as 'nomy', an appropriation and reversal of its antonym, 'anomy', a standard term in Sociology;
2. The composition of the corpora, related to the language modality (oral and written), of the types and textual genres and the selection that could have been done on the total volume of normative signs due to rules of occurrence, frequency or specialized use;
3. The methodological aspects adopted for the macro and micro-structural treatment, especially in terms of normative rules from the concept of variant, if it is obviously the scope of the work, and according to the degree of complexity of the remission system;
4. The size and degree of accuracy that is aimed for in the semantic field of each lexical unit to be observed, that is, how large the surveys will be on the various pos-

17 Translation: specialized language is what a linguistic subsystem that combines linguistic specificities of a particular area is called. In fact, the terminology, if considered in its original concept, highlights the concepts and terms of their own domain. From this angle, it is too much to talk about a specialized language, it being more appropriate to refer to this as specialized vocabulary.

18 Although some authors argue that the "lexeme" would be the unit of lexicological and lexicographical treatment, it is believed that the word *lexie*, proposed by Pottier (1974), is more appropriate scientifically, since it allows longer or shorter units of what is usually known as "word" to be taken within the scope of investigation, presupposing that the *lexie* is the result of agglutination or composition of n-lexemes with n-grammemes, at the moment of interaction at the lexical level between signifier and signified.

sible meanings that the word can take within its contextual uses, if this is of interest for the research.

Depending on the degree to which each of these aspects is adopted, it will be possible to establish clear distinctions between the products of lexicographical or terminological work such as glossaries, dictionaries and vocabularies, since although they are scientific products, each should require epistemological and methodological decisions that are grounded in these and in other theoretical guidelines that can be complementary.

Furthermore, the investigation of the historical lexicon is necessarily dependent on the dialogue with those different approaches and subject to the varying notions of otherness, because otherwise it leads to great inaccuracies. The permeabilization of the history of the lexicon of a language should attempt to break disciplinary boundaries, moving towards an object of a heterogeneous and multifaceted nature that is human language, and one of its main structures: vocabulary.

4 The lexical constitution of Brazilian Portuguese

4.1 About the first written records and lexical dynamicity

If one were to see some of the first documents written on the Atlantic shores of the southern West—an example of the awakening of Brazil's history and the lexical inventory of Brazilian Portuguese—one would be able to perceive how the lexicon of seafarers, reflected in Caminha's Letter here, was able to convey, through the awakening of the story, the overall assets of the language. Let us take the following excerpt of this beautiful work as an example:

> foram xxjdias dabril que topam*os* algũũs synaaes de tera seemdo da dita jlha seg*und*o os pilotos deziam obra de bj^{c} lx ou lxx legoas os quaaes herã muita camtidade deruas compridas aque os mareantes chamã *botelhoe* asy outras aque tam bem chamã*Rabo dasno*. Eaaquarta feira segujmte pola manhãã topamos aves aque chamã *fura buchos e*neeste dia aoras de bespera ouuem*os* vista de tera (Caminha's Letter, fólio 1r).[19]

Analysing the term *botelho*, we can verify that this lexical element is today part of the common lexicon, occurring in Portuguese dictionaries, and constitutes an elegant result of homonymic convergence. In this way, different original etymological bases be-

19 Our reading of the original manuscript. A rather free translation would be: 'On the 21st of April, being about 660 or 670 leagues away from the aforementioned island, according to our pilots, we found several signs of land, namely a large quantity of long herbs, which the seamen call *botelho*, and also others, which they call *rabo d'asno*. And the next Wednesday in the morning we encountered birds which they call *furabuchos*. And on that same day in the evening we got sight of land'.

gan a process of metaplasm, causing it to become a single phonic form and sometimes a graphical one, at least for a particular period in history or in some particular dialectal space, until a new set of changes began to occur.

In other words, the *botelho* of Cabral's sailors, a word which in that context referred to a generic name for what is known as *carvalhinho do mar*, a type of sea plant, from the root word *abutilón*, in Castilian, lived concurrently in Portuguese with other units with distinct semantic loads or from other etymological sources, yet with the same signifier.

This means that at the same time that Caminha's Letter was written, Portuguese had the homonymic element *botelho*, which referred to a unit of measurement smaller than the *celamim* (2.27 l.) and whose etymological origin is the French *bouteille*. *Botelho* and its marked form *botelha* would already have been variants when, in the first half of the 17th century, Nuno Alvares Botelho, acting viceroy of India, lent his name to a currency that he had introduced, but it did not escape the same processes of phonic change to which the namesake pair here underwent.

As history developed further, the forms *botelho* and *botelha* again underwent new metaplasms, a trend of secular drift of the sonorization of intervocalic stops. In this way, they also started to display the variants *bodelhos* and *bodelhas*, which in turn entered the common lexicon. According to Viaro (2011, 172) 'the transformation of the originally voiceless consonants into voiced ones, in the intervocalic position, is a phenomenon common to all romance languages and dialects.'[20]

If any speaker of Portuguese has the curiosity to consult a dictionary, they will find that there are different entries for *botelha, botelho*, *bodelho* and *bodelha* because, from a lexicographical perspective, they are firstly convergent forms, and then divergent ones. They also possess different meanings, which cannot be characterised in a singular way within their semantic fields. Machado Filho (2014) offers a simple example of the phenomenon of convergence from a known case of homonymy:

> 'Perhaps the lexie *manga* is the best known example, in Brazil, of a convergence related to the historical confrontation between hereditary change and loan, as 'part of the clothing that covers arms' and its significant extensions by metaphor through a Latin root, *manica*, and what the 'pulpy fruit with central core, large and full of pith' of Malaysian etymology represents' (Machado Filho 2014, 269).[21]

But it is known that the denominative role in a language does not always depend on the lexicon, since other strategies may emanate from different levels in the process of

20 "a transformação das consoantes originalmente surdas em sonoras, na posição intervocálica, é fenômeno comum a todas as línguas e dialetos românicos."

21 "Talvez seja a lexia *manga* o mais conhecido exemplo, no Brasil, de convergência relacionada ao confronto histórico entre mudança hereditária e empréstimo, já que 'a parte do vestuário que recobre os braços' e suas extensões significativas por metáfora são de étimo latino, *manica*, e o que representa o 'fruto polpudo com caroço central, grande e fiapento', de etimologia malaia".

meaning construction. Lexical items such as *botelho*, *rabos d'asno* and *fura-buchos* were grouped together in the same folio collection of Caminha, a morphosyntactic strategy to express objects whose lexicon was still not properly manifested in Portuguese or had yet to be absorbed as a loan, such as the case of *cocar*,[22] a gift offered by the Indians to the captain on their first contact:

> E hũũ deles lhe deu hũũ *sombreiro de penas daues compridas cõ hũũa copezinha pequena de penas vermelhas epardas coma de papagayoe* outro lhe deu hũũ Ramal grande de comtinhas brancas meudas que querem pareçerdaljaueiraas quaaes peças creo queo capitam ~~n~~ manda avossa alteza (Carta de Caminha, fólios 1v e 2r).[23]

These examples reinforce the understanding of the dynamics that the process of creating, importing and renewing the lexicon of a language undergoes. They show us that "under the skin of the words" there is much more than "ciphers and codes", if one thus dares to disagree with and hence overcome (from a sociolinguistic perspective) the very well-known verses of Drummond de Andrade ([30]2005, 27), the poet with whom we began this article.

4.2 Language contact and lexical borrowings

Brazilian Portuguese would, from the first contacts, come to experience the plural effects that socio-history imposes on all natural languages.

With the arrival of the colonisers for the implementation of the exploratory company of "overseas", new lexical units began to be absorbed into the inventory of Portuguese. This happened in different times during the more than 500 years of history in which such influences on vocabulary were at their most intense, especially from indigenous people, like the Tupi linguistic branch, and African people, in particular and in greatest volume the Bantu linguistic subgroup, and later the lexical waves of immigrants during the 19th and 20th centuries.

For a chronology of the historical facts from this period, ↗2 The social history of Brazilian Portuguese and ↗3 The history of linguistic contact underlying Brazilian Portuguese. Nevertheless, it is important to note that colonisation begins between 1532 and 1534. In 1549, the first general government of Bahia was established. From there, the Portuguese dominated the native peoples, indeed with many of them now being exterminated. Before that, however, a general language based on Tupinambá

22 The most curious thing to note is that *cocar* is of French etymology, *cocarde*, so it has expanded into Portuguese. The Tupi term would be *acangatara*.

23 Our reading and highlights. A rather free translation would be: 'And one of them gave him a sombrero of long bird's feathers with a small cup of read and brown feathers, like those of parrots. Another one gave him a long braid of little white *comtinhas* that look like little pearls. These pieces, I think, are sent to you by the captain'.

developed, as well as the beginning of the slave trade, which would last until 1850 when the Eusébio de Queiroz law was passed.

Coelho (2000, 57), when talking about Portuguese navigation, trade and conquests, says:

> 'the words that in this century [15th] started to designate this prodigious collective movement were discoveries, expansion, evangelisation, empire, civilisation encounters, dialectic of the other and the same, to civilise, slavery, colonialism, building new nations and countries, time of discovery of the naked and shame, passing from the particular to the universal'.[24]

The author here in fact references so many words that quoting them in full would be unfeasible. It is interesting to note that a generous lexical inventory is aggregated to these Portuguese words as a dialectical response of indigenous languages to the contact imposed by the Portuguese, like "les forces de résistance à la glottophagie"[25] (Calvet 1974, 79).

4.2.1 The indigenous languages contribution

According to Ribeiro (1997, 98), "[I]ndigenous slavery prevailed throughout the first century. Only in the 17th century would [B]lack slavery overwhelm it".[26] But from the languages of these periods of slavery contact, "no other influenced Portuguese so much as Tupi" (Viaro 2004, 282).[27]

Portuguese today includes around a thousand inherited lexical items from Arabic, inherited during the Islamic rule on the Iberian Peninsula, which occurred between 711 and 1249 in Portugal, and around five hundred inherited words from Germanic languages. But it will not be possible to make a numerical count, on the same lines, of the effective contribution of Brazilian indigenous languages, without the risk of great uncertainty, until a systematic and historical research in language dictionaries and vast corpora is performed, although it is suspected that it exceeds the Arab heritage, at least in the passive vocabulary.

Nascentes (1951), in the composition of his etymological dictionary of the Portuguese language, says:

---

24 "os vocábulos que neste século [XV] passaram a designar esse prodigioso movimento coletivo foram descobrimentos, expansão, evangelização, império, encontro de civilizações, dialética do outro e do mesmo, civilizar, esclavagismo, colonialismo, construção de novas nações e países, tempo da descoberta do nu e das vergonhas, passagem do particular ao universal."

25 'the forces of resistance to glottophagy'.

26 "A escravidão indígena predominou ao longo de todo o primeiro século. Só no século xvii a escravidão negra viria a sobrepujá-la."

27 "nenhuma outra influenciou tanto o português quanto o tupi."

> 'For Tupi, we have the works made by the Jesuits interested in the catechesis of Indigenous people in the early days of discovery and conquest. We also have the works made in the last century and the present. But despite all of these, the documentation is still precarious' (Nascentes 1951, x).[28]

He goes on to say that 'Tupi is not made on the asphalt. It is made in the jungle, in contact with the Indigenous people' (Nascentes 1951, xi). Nascentes thinks that 'modern works, despite the jingoism, are pure works of dilettantism, representing great effort, a lot of goodwill, but of dubious scientific value' (Nascentes 1951, x).[29]

In 1978 Antônio Geraldo da Cunha published the *Dicionário histórico das palavras portuguesas de origem tupi*,[30] with several subsequent editions. It remains one of the major references in the field. Prefaced by Antônio Houaiss, the dictionary records the chronological variants, with greater emphasis on texts of the sixteenth and 17th centuries, and less so on those from the 18th to 20th centuries, from those that constituted its nomenclature. It also provides an index of all variants identified in the corpora with around 5,030 lemmatic signs, i.e., the graphic-phonic and morphological variations which, within the lexicographical work, could be interpreted as secondary and as multiple lemmas, depending on the desired reference system.

It is possible to cover almost the entire alphabet in the exemplification of Tupi etymology in Portuguese language: *abacaxi*, *aipim*, *baiacu*, *buriti*, *caatinga*, *cajá*, *caju*, *capim*, *capivara*, *carnaúba*, *cipó*, *cupim*, *curumim*, *emboaba*, *enxu*, *gambá*, *guri*, *igapó*, *igarapé*, *imbu*, *ingá*, *jaburandi*, *jaburu*, *jabuticaba*, *jacarandá*, *jacaré*, *jandaia*, *lambari*, *macaxeira*, *mandacaru*, *mandioca*, *mangaba*, *mingau*, *muriçoca*, *nhambi*, *oca*, *oitizeiro*, *paca*, *peteca*, *piaba*, *quiriri*, *quixabeira*, *sabiá*, *saúva*, *siri*, *suçuarana*, *sururu*, *tabaréu*, *taboca*, *taquara*, *tatu*, *urubu*, *vitinga*, *xexéu*, *zabucai*.

However, hardly any lexical item that starts with a voiced dental occlusive phoneme, represented by the letter <d>, will be from Tupi etymology, except when it is a derived form in Portuguese, such as *descoivarado*, also presented by Cunha (1978) as the only occurrence in the *corpus*. Of course, any random list will show how indigenous contributions are heavily concentrated on toponyms and lexical elements mainly related to flora and fauna.

Viaro (2004, 282–292) presents a list of 169 lexemes, extracted from the previously mentioned dictionary of Cunha, where 25 % of the lexemes refer to national fauna and 18 % to national flora. As the largest group, 28 % of the items on this list refer to place names, confirming what Calvet says (1974):

---

28 "Para o tupi dispomos dos trabalhos feitos nos primeiros tempos do descobrimento e conquista, pelos jesuítas interessados na catequese dos indígenas. Dispomos também de trabalhos feitos no século passado e no atual. Mas, apesar de todos eles, ainda é precária a documentação".

29 "trabalhos modernos, em que pese ao ufanismo, são puras obras de diletantismo, representativas de grandes esforços, de muita boa vontade, mas de valor científico duvidoso".

30 Historical dictionary of Portuguese words stemming from Tupi.

> la toponymie est sans doute le substrat le plus résistant aux strates successives de langues qui se succèdent, se remplacent ou se déglutissent en un point particulier du globe[31] (Calvet 1974, 105).

In a way, Ramos/Venâncio (2002) confirm this idea by observing toponyms from Minas Gerais in order to verify "intellectual or religious fashions and political sensitivity variations" (Ramos/Venâncio 2002, 115) as a source for the writing of the social history of the Portuguese language in Brazil. The authors conclude that "in the last hundred and fifty years, 9.9 % of Minas Gerais localities have adopted indigenous names" (Ramos/Venâncio 2002, 116). The authors present the names of locations from Minas Gerais, considering in their work the period from 1711 to 2001. The adoption of indigenous place names represents 45 % of the data, and is followed closely by the use of personal names, at 38.1 %, revealing that anthroponymy was also quite significant (↗18 Onomastics and toponomastics).

Recently, Coutinho (2010) has collected, in a single study, words from different languages that could, from his perspective, have contributed to the formation of the Brazilian Portuguese lexicon. He found them among the African languages, as well as Arabic, English, French, Italian and Tupi. For the latter, he organizes the units on onomasiological lines into 26 conceptual fields or subject areas, the largest of these being flora, followed by ethnology. In his study, strangely, place names were not considered.

For flora he registers, among many other items: *açaí*, *andiroba*, *babaçu*, *butiti*, *cabuçu*, *embaúba*, *guabiroba*, *guarajuba*, *indaiá*, *jaborandi*, *macambira*, *oiti, pequi*, *samambaia*, *taioba*, *ubatã*, *vuarame*, *xuru* etc.

4.2.2 The impact of African languages on the lexicon

Regarding the contributions of African languages to the composition of the Portuguese lexical inventory, it must be considered initially that speakers of a single language family—the *Niger-Congo family*—this from the *Congo-Cordofanian* branch, would have come to Brazil throughout the slavery period. A small group of Islamised speakers of the Hausa language, another linguistic branch—the Afro-Asian—would come in the 19th century to a few urban centres, without having a large lexical impact on southern Portuguese on this side of the globe.

From the Niger-Congo family, the most commonly represented languages were those of the Bantu subgroup (Kimbundu, Kikongo and Umbundu), all of the Benue-Congo branch, and Yoruba languages (also known as Nago), Ewe, Fon, Mahi (or Jeje

31 'toponymy is, undoubtedly, the most resistant substrate to the successive strata of the following languages. They supplant or devour each other in a specific part of the globe'.

in Bahia) and Ijo, from the Kwa branch. This ethno-linguistic composition would remain in Brazil due to the early adoption of slavery by the Portuguese colony.

Going back in time, in search of the presence of people of African origin in Portugal at the beginning of the great navigations, Tinhorão (1997, 86) presents a table of the slave trade between 1441 and 1505, estimating that between 136,000 and 151,000 people were brought to this country in that short period of time. Tinhorão believes that the most perennial lexical loans to the Portuguese language would have been those of Bantu origin, including a small glossary with 94 lexical entries, of which 39 (41 %) were related etymologically to Kimbundu.

Similarly, and looking at the well-known work of Mendonça (1948), who came to conduct a survey of vocabulary of an African origin used by Brazilian writers, the record of 393 words allegedly from African origin is found also in Brazil. From this list, and bearing in mind that the claims of etymological research are not always accurate, about 34 % of the words would be based on Kimbundo etymology and 13 % on Yoruba. However, many of these are labelled with the generic designation of "African term". Elements of Arab origin, transmitted more recently through African speakers during the colonial period, are also identified by the author.

According to Pessoa de Castro ([2]2005, 34), among African speakers of Bantu languages, 'the coastal people were notable for numerical superiority, duration and continuity in direct contact time with the Portuguese colonisers: 1) *Bacongo*, 2) *Umbundu* and 3) *Ovibundo*'.[32] Kimbundu was the language of the Mbundu, while Kikongo and Umbundu were spoken by the Bacongos and Ovibundos respectively. Data from Mussa (1991) confirm this information and note the Mbundu as a demographic exponent during the slavery period in Brazil and Kimbundu as the majority language.

Medeiros (2008) recently carried out a detailed count of the lexical items considered to be African loans into Brazilian Portuguese as present in one of the most widespread electronic dictionaries of Brazilian Portuguese, the *Aurélio Século XXI* (see Ferreira 1999). 781 records described as being of African etymology were found. Of these, 156 were judged to have inaccurate or vague etymological records, being identified as mere Africanisms. From the initial 781, approximately 37 % were from Kimbundu; 29 % from Yoruba (Nago) and only 5 items from Hausa and other languages.

Although Medeiros (2008) thus finds an unexpectedly low number in comparison to the 2,500 items identified by Schneider (1991) identified,[33] who used the same basic corpus, Medeiros notes that

> 'in the dictionaries mentioned, there is no record of lexical items with the etymological classification of 'Europeanism'—it would not make sense, since there are many European languages—and those African languages are also many—and families and branches are fairly disparate [...]

32 "[...] destacaram-se pela superioridade númerica, duração e continuidade no tempo de contato direto com o colonizador português, três povos litorâneos: 1) *bacongo*, 2) *ambundo* e 3) *ovibundo*".

33 The difference between the results of the two authors deserves in-depth research.

> —there is no logic in adopting the label 'Africanism' as an etymological classification' (Medeiros 2008, 15).[34]

This position of Medeiros, which, after all, becomes an affirmative action, is extremely important, since much of the records of linguistic history have been lost due to the poor visibility given to the role of minority languages that were in contact with socially and economically dominating languages. Africa is not a country but an entire continent in which the language issue should be given the same treatment as that of so-called "historical or cultural languages", whose relationship with writing and with military and economic power meant that they have become "victorious", as Serafim da Silva Neto (1960, 25) used to say, 'for a standard language [...] was overlapping the others like a progressive oil slick'.[35]

Gonçalves/Chimbutane (2009) believe that

> 'the studies that have addressed the linguistic contact situations of formerly colonised regions have not valued particularly the different roles and statutes of languages in contact, not taking into account that because they occupy very different places in the communities in which they coexist, local languages and former colonial languages do not have the same mutual impact' (Gonçalves/Chimbutane 2009, 31).[36]

What is surprising, however, is that despite this asymmetry, especially if one considers that African languages did not have contact with the relative permanence afforded by written registers, the African lexical legacy in Brazilian Portuguese is quite substantial.

Bonvini (2002) observes that, very early, in *Asia*—a monumental work by João de Barros—African lexical elements began to be uttered, in different *decades* (the term adopted for each of the main parts of his work), like *banzeiro*, *fulo*, *furna*, *inhame* (Bonvini 2002, 149), among others. These same elements arrived in Brazil, and some remain in the active vocabulary of different speech communities.

Alkmim/Petter (2008) conducted very influential research into lexical items currently used in Brazil, looking also at the historicity of these. Basing their study on the list of words recorded by Viscount of Pedra Branca on lexical differences between Portugal and Brazil, the authors show that in the period around independence, 16 words

---

34 "nos dicionários citados, não há registro de itens lexicais com classificação etimológica do tipo 'europeísmo'—nem faria sentido, uma vez que muitas são as línguas européias—, e que as línguas africanas também são muitas—e de famílias e ramos bastante díspares [...] —, não há lógica na adoção do rótulo 'africanismo' para classificação etimológica".

35 "[...] a linguagem padrão [...] se foi sobrepondo às demais, como progressiva mancha de azeite".

36 "os estudos que se ocuparam das situações de contacto linguístico das regiões outrora colonizadas não têm valorizado particularmente os diferentes papéis e estatutos das línguas em contacto, não tomando em consideração que, pelo facto de ocuparem lugares muito distintos nas comunidades em que coexistem, as línguas locais e as línguas ex-coloniais não têm o mesmo impacto mútuo".

of African origin, unknown in Portugal, were in normal use in Brazil: *batuque*, *caçula*, *cochilar*, *fuxicar*, *mandinga*, *mocotó*, *mulambo*, *mungangas*, *muxiba*, *muxiqueiro*, *muxoxo*, *quindim*, *quitanda*, *quitute*, *xingar* e *senzala*. They then note that in the period following independence and until the early 20th century, research under the label Africanism were, unfortunately, not systematic.

Meanwhile, research on the influence of African languages in terms of the lexical composition of Brazilian Portuguese has surfaced once more from the 20th century onwards, and the new studies will expand

> 'just over a hundred lexical items recorded in the 19th century [...] to more than 300 in the first half of the 20th century, exceeding 2,000 items in specialized dictionaries, published in the late twentieth century' (Alkmim/Petter 2008, 150).[37]

This points to a very important issue for lexicon studies from a historical perspective: the notion of active and passive vocabulary of speech communities in regard to the general assets of the language. The active vocabulary, as a rule, is associated with the immediate or automatic encoding and decoding process by speakers of lexical data in current linguistic use. Active vocabulary, however, is not the same as the notion of lexicon in circulation, because the former is a subsystem of the latter.

Lexical heritage, however, cannot be measured solely through the number of items. It has to be done according to the degree of the maintenance of the historical transmission of vocabulary and the frequency of its use. One area of lexical studies, lexical statistics, has focused on precisely these issues. An example here is the study by Biderman ([2]2001), which conducted a textual analysis of some 5,000,000 words, the composition of which covered the period 1950–1995, and concluded that the tabulation of the data allowed him to state that "80 % of any Portuguese text is made up of [...] 1,000 words, which are repeated continually" (Biderman [2]2001, 336). Obviously, for lexical studies, content words are more interesting, i.e., nouns, adjectives and verbs, and these are commonly called lexical words.

Returning to the lexical contribution of African languages, the major significance of Alkmim/Petter's research is that they sought to follow the principles of real use and the vitality of the lexicon. To test the availability of some lexical items of African origin to be decoded, the authors selected 249 words, 400 originally identified by speech communities, and looked at the degree of familiarity of these items for 125 informants. These speakers were able to validate 56 words, grouping them into three categories: i) elements used in any context of social interaction; ii) elements used colloquially in any way; iii) and elements with informal and restricted uses.

---

37 "pouco mais de uma centena de itens lexicais, registrados no século XIX, [...] a mais de 300 na primeira metade do século XX, ultrapassando 2000 itens nos dicionários especializados, publicados no final do século XX".

In the first category we find *abadá*, *banzo*, *caçamba*, *cachaça*, *cachimbo*, *caçula*, *candango*, *canga*, *capanga*, *carimbo*, *caxumba*, *cochilar*, *corcunda*, *dengo*, *fubá*, *gibi*, *macaco*, *maconha*, *macumba*, *marimbondo*, *miçanga*, *molambo*, *moleque*, *moringa*, *quilombo*, *quitanda*, *quitute*, *senzala*, *tanga* and the verb *xingar*.

In the second category, *bamba*, *bambambã*, *banguela*, *cafuné*, *catimba*/*catimbeiro*, *catinga*, *mandinga*, *muamba* and *muxoxo*.

Finally, the third category included *angu*, *babaca*, *babau*, *biboca*, *bunda*, *cafofo*, *cafundó*, *cambada*, *cucuia*, *muquifo*, *muquirana*, *muvuca*, *muxiba*, *quizumba*, *sacana*, *ziquizira* and *zumbi*.

Regardless of the textual records in which the above elements were used, it is true to say that, in general, all 56 lexical units here are easily recognized by Brazilians nowadays, although it is probable that many of these speakers have absolutely no idea that these words are etymologically related to one or other African language. Incidentally, the mandatory teaching of "History and Afro-Brazilian Culture", established by Law No. 10,639, 2003, has not yet been rolled out effectively in the country. Etymological studies have also rarely been a concern for Brazilian schools, and this in turn makes the population generally very unaware of their socio-historical heritage, especially the cultural and linguistic contacts to which the Portuguese language was exposed during the process of its formation and constitution.

4.2.3 The immigration influence

Apart from the contribution of indigenous and African languages, it is important to understand the migration processes in Brazil and their impacts from a linguistic point of view, especially regarding lexical renewal, which may have occurred mainly with language loans from Europe and Asia, although Ribeiro (1997, 242) believes that the immigration influence, while 'significant to the racial and cultural constitution [of the southern areas], had no more relevance in determining the characteristics of Brazilian population and its culture'.[38]

Ribeiro (1997, 242) notes that in the period 1851–1970, the "integrated European immigration contingent in the Brazilian population [should be] valued at 5 million people."

This information can be confirmed with data published by the Brazilian Institute of Geography and Statistics (IBGE 2000, 266). The total migration flows occurring in the following decades are registered thus: 1884–1893 (883,668); 1894–1903 (852,110); 1904–1913 (1,006,617); 1914–1923 (503,981); 1924–1933 (717,223); 1945–1959

38 "[...] relevante na constituição racial e cultural [das áreas sulinas], não teve maior relevância na fixação das características da população brasileira e da sua cultura".

(667,094), and this influx was composed mainly of Italians, from 1884 to 1903, and after that, people from Portugal.

This can be compared with the evolution of the overall population, which in 1850 was somewhere around eight million, growing to 9,930,478 by 1872 (the year of the first Brazilian census), to 17,438,434 by 1900, and to 30,635,605 by 1920. Considering that migratory communities tend to be concentrated in previously established communities, it is necessary to bear in mind that the impact that immigrant languages may have been quite significant, contrary to what Ribeiro (1997) believed. Although restricted to some regions, these languages were in a strong position for potential national expansion, depending on the prestige of the emanating centre.

The matter of linguistic prestige is extremely significant to the consolidation of linguistic norms and, in particular, for the shaping of the lexicon of a language and, consequently, for the adoption of loanwords.

Garcez/Zilles (2002) consider that

> 'different groups in a community can assign different values to the identities linked to speakers of other languages. The values associated with foreignness can often then come into conflict within the community that makes the loan' (Garcez/Zilles 2002, 15).[39]

Prestige can be measured exogenously, in international terms, as was the case of French, which largely dominated the expansion of loans in the modern world for centuries, especially during the period of the *Belle Époque*—between the last quarter of the 19th century and the first 15 years of the 20th century—leaving mark on many Western languages, because of the importance that was historically given to France and this period in particular in the Western world.

For different reasons, and engendered by economic power and technological innovation, English subsequently began to assume the role of the emanating generator of loans, including in Brazil.

Barcellos da Silva (2000, 145) says:

> 'Change is inherent in all living languages and in the case of technology-importing countries, such as Brazil, it is inevitable that scientific and technological vocabulary, and even the ordinary one, will implement words derived from where the new meanings emigrated'.[40]

In the case of immigration languages in Brazil, it would be at least naive or even a "linguistic prejudice to assume that what is perceived as foreign words today will remain for a long time something strange and alien to the receptor language element,

39 "diferentes grupos em uma comunidade podem atribuir valores diversos às identidades ligadas aos falantes de outras línguas. Então os valores associados a um estrangeirismo podem muitas vezes ser conflitantes dentro da comunidade que faz o empréstimo".

40 "A mudança é inerente a todas as línguas vivas e, no caso, de países importadores de tecnologia, como é o caso do Brasil, é inevitável que o vocabulário científico e tecnológico e mesmo o corriqueiro, de cada dia, seja implementado por palavras oriundas das noções donde as novidades emigraram".

keeping its alien load" (Garcez/Zilles 2002, 33), since, concurrently with the lexicon importing process, the forces of the linguistic system are acting towards the phonetic, morphological and morphosyntactic accommodation of foreign innovations. One example is the word *delete*, which originated from English *to delete*, and probably entered Brazilian Portuguese through writing, in that it no longer expresses the phonic characteristics of its etymology, behaving morphologically in compliance with the standard Portuguese and its predilection for the first conjugation with *-ar*.

If we consider there to be no linguistic original genesis, but rather continuous and historical processes of constitution, all these merged by social, economic and political contingencies of contacts that cross each other's paths, it would be really absurd to accept the Bill No. 1676/99 as defended by one congressman, especially in its Article 4, which proposed considering "any use of foreign word or expression [...] harmful to the Brazilian cultural heritage." John Ribeiro (apud Cardoso/Cunha 1978, 249), in 1933, had already noted, in relation to the dynamics of the national language, "that our Independence still maintains these handcuffs on the wrists, and that the American personality pays tribute to the submission of words".

The truth is that, to a greater or lesser extent, the languages of immigration have left their mark on the lexicon of Brazilian Portuguese, not only as a direct etymological heritage, namely the transmission of certain words of these languages by contact, whose units settled into the phonological pattern of the Portuguese language, but also through indirect contribution, i.e. the adoption of pre-consolidated loans in the source language, acquired from other languages in previous exchanging situations.

The school dictionary of the Portuguese language (FENAME 1956) features in its 1978 edition (FENAME 1978) an appendix with the inclusion of Latin and foreign expressions, including words and phrases in French, English, Italian, and Japanese.

It is interesting to turn to this dictionary here, because it is an official publication funded by FENAME, the National Fund for School Supplies, whose actions are based on the National Textbook Program (PNLD). The registration of lexical and sentence units ensures the idea that the so-called foreign words, especially in the cases of immigration language loans, would have taken more stabilized linguistic values in the lexicon of Brazilian Portuguese.

However, it would be hard to decide whether the listed elements integrate historical cultural contact processes or effective language contact that has already occurred in Brazil. This demonstrates the importance of some ongoing studies into etymology in the country, like the retrodating project of the Portuguese lexicon conducted by NEHiLP Group, from USP, referred to above. From the Italian language, the following items are found in this list: *batuta*, *caricatura*, *ferrovia*, *fiasco*, *imbróglio*, *lasanha*, *maestro*, *mortadela*, *novela*, *palafita*, *pérgola*, *ravioli*, *salame*, *solfejo*, *tômbola*, *torta*, *viola*, *violin* and *virtuoso*, among others.

The list of lexical units stemming from Japanese is much more modest, including *biombo*, *caratê*, *gueixa*, *judô*, *quimono* and *samurai*. In this list, besides English and French, other languages were not included, although they were spoken by significant

contingents of immigrants who arrived in Brazil. For example, languages like Spanish, German, Syrian, Turkish etc.

The French and English elements are obviously a reflection of the cultural contact discussed previously, and they spread through different conceptual fields in Brazilian Portuguese. From French, Brazilian speakers of secondary education often recognize *balé*, *bijuteria*, *bufê*, *butique*, *cachecol*, *cachepô*, *chance*, *dossiê*, *garçom*, *grand prix*, *maître*, *ménage*, *omeleta ~ omelete*, *revanche*, *réveillon* and *sabotage*, some of which were never adapted in writing, as can be perceived.

As for English, any list would most likely be incomplete without it, due to the volume of anglicisms in colloquial vocabulary. Nevertheless, in favor of the historical record to which any text can be repairable, and especially because some elements do no evince their etymology immediately, this due to the degree of adaptation of the phonic that underwent, we will cite here *baby-doll*, *babylook*, *bangalô*, *bangue-bangue*, *bonde*, *box*, *checape*, *check-in*, *drope*, *ferry-boat*, *futebol*, *gay*, *hall*, *hambúrguer*, *hobby*, *jeans*, *jipe*, *miss*, *piquenique*, *náilon*, *o.k.*, *tanque*, *tíquete* and *w.c.*, among many others.

4.3 Neologisms

Moving from the concept of loan to that of neologism, we can appreciate how this latter includes, in addition to incorporating items from other languages, the so-called loan neology creation processes facilitated by our language system, especially its morphology and morphosyntax, operating in accordance with the creative din, or formal neology.

Guimarães Rosa (1967, 64) says that 'a neologism bruises, confuses, almost offends'. He also warns that 'if everybody goes off into the world to find their words, where do you go to with the current and inherited language?'[41]

The notions of system, standards and speech are certainly well settled for Guimarães Rosa. The lexical creation in any linguistic variety is a solitary act, and therefore an act of speech, in which the individual makes use of the system's possibilities to imprint, in a signifier, the meaning aimed at for a representation. A parallel of this may be established with music.

The sounds of a possible range of notes can be organized for the composition of melody lines, chords or indeed a symphony. For each composer, the result will never be the same, because the possibilities of articulation of these sounds and their harmonies are endless and subject, in addition to rhythm and pauses, to the creative process of the author and his or her artistic genius. Nevertheless, if a melodic verse, song or symphony, once created, is accepted by those who hear it, then it can be reiterated re-

41 "um neologismo contunde, confunde, quase ofende"; "saia todo-o-mundo a empinar vocábulos seus, e aonde é que se vai dar com a língua tida e herdada?"

peatedly, consolidating its position as a cultural habit, like a model. This corresponds broadly with what is going on with language.

Neology, then, follows this pattern, and the possible outcomes arising from its manifestation are the following:

i) Merely formal neology, that is, the engineered unit, is not supported by the language communities to which it was exposed and the alleged lexical element will be nothing but a speech act;
ii) Lexical innovation: the form is repeated and maintained by other speakers of the language and can be restricted to a particular community, which might later contribute to the characterisation of speech or a dialect, or even expand to other diatopic or diastratic spaces of the language;
iii) The lexical unit, after a long period of use, falls into disuse for one reason or another, becoming a member of the lexical assets of the language, and condemned as archaic in dictionaries until, perhaps, "wind's hand castill lift it",[42] as Fernando Pessoa (1977, 83) would say;
iv) After being created and used in the language, a unit can win or lose significant traces or semes, causing its original integrity to undergo significant changes, even deep ones, either in neology, in terms of semantics, or in its conceptual sense.

Guilbert (1973) considers

> néologie sémantique tout changement de sens qui se produit dans l'un des trois aspects signifiants du lexème sans qu'intervienne concurremment un changement dans la forme signifiante de ce lexème. La première forme de néologie sémantique est celle qui s'opère dans le changement du groupement des sèmes afférents à un lexème, selon des modalités diverses. Celles-ci ont été décrites par les rhétoriciens sous le nom de synecdoque, métaphore, comparaison, métonymie (Guilbert 1973, 21–22).[43]

On this point, a significant issue in the formation of the Brazilian Portuguese lexicon is the behavior of lexical items themselves. According to Barbosa (2000),

> 'every new proposition of a sign implies not only the production of a new anthropo-cultural frame and linguistic unit corresponding to it, but also the response to those needs, in a wider context. In this sense, the mechanisms of the formation of new signs, or the assigning of new meanings to ex-

42 The original spelling is followed here, as in *Message*, the only book published by the author during his lifetime.

43 '[...] semantic neology all the changes in sense which are produced in one of the three significant aspects of the lexeme without intervention of a concomitant change in the significant form of this lexeme. The first form of semantic neology is the one which occurs in the change of the sememe of a lexeme, in accordance with several modalities, which are described by rhetoricians as synecdoche, metaphor, comparison, and metonymy'.

> isting signs, often present themselves as complex processes, the formulation and selection of proposals made within the social group involved' (Barbosa 2000, 177).[44]

In other words, the complexity of the change of meaning of a given lexical item conforms to the same force and the same conjunctive and disjunctive spectrum operating in a functional system, and at the same time to the extra-linguistic factors that are imposed on the language as a whole in the different levels of analysis, and which are grounded in varying rules. Perhaps the full maintenance of a meaning over the course of history might be considered more surprising than the change of semic load of a particular lexical item, due to the constant pressures caused by the emergence of variants, especially in irregular linguistic transmission situations, as was the case in the historical transplantation of Portuguese to Brazil.

The history of Portuguese is full of semantic neologisms and any list would be incomplete here. However, an excellent example is the verb *falecer*, which was used in the Portuguese of the archaic period long before the Portuguese landed in Brazil, with the etymological sense of 'missing', 'exhausting', inherited from Latin **fallescere*, as illustrated in the excerpt from a medieval *Flos Sanctorum*, "entrou no celeiro e vyo que o pã *falecia*" (Machado Filho 2009, 227), i.e., 'he went to the cellar and saw that bread was lacking'; however, in our times the word's semantic load is restricted to 'die' in all the language varieties, at least as far as we know.

This trend is manifested through the obeying of lexical dynamics, and Brazilian Portuguese is full of semantic neologisms. A glance at the *Linguistic Atlas of Brazil* (ALiB) published in 2014, shows how social group is one of the determining factors for certain changes in the semic values of lexical units. By observing, for example, the different meanings registered in the country's capitals for *prostituta*, we can see that one of the significant variants, *rapariga*, is present in all the capitals of the Northeast, in addition to Belo Horizonte, Cuiabá, Rio Branco and Boa Vista (cf. Cardoso et al. 2014b, 229), although in Portugal today this item has the value of 'child or young female'.

5 Unity and diversity in the so-called Lusophone world

This brings us to the discussion of the unity of ideas, diversity, and unity in diversity, which in the area of the lexicon are very complex ones, in that phonic and lexical aspects are typically seen in Geolinguistics as the beacons to be used in establishing dialectal boundaries:

---

44 "cada nova proposição de signo implica não apenas a produção de um novo recorte antropocultural e da unidade linguística que lhe corresponde, como também a resposta àquelas necessidades, numa conjuntura mais ampla. Nesse sentido, os mecanismos de formação de novos signos, ou de atribuição de novos significados a signos preexistentes, apresentam-se, frequentemente, como processos complexos, de formulação e de seleção de proposições feitas no seio do grupo social envolvido".

> 'In this intersection of social and spatial dialects and linguistic standards, specific standards of written language that neutralise many of the differences of everyday speech intervene, but are far from being dismissed. The backbone, however, that unites all these differences is captured at a higher level of abstraction, that is, the system of common rules underlying these differences, and which supports the idea that, as a historical phenomenon, it can be said that in such different points of the world, the Portuguese language exists, and not another language' (Mattos e Silva 1988, 13).[45]

This is the same as saying that the lexicon is not the best scenario to discuss *language* as a whole system, but maybe we can talk about *language norms* and, therefore, the spatial distance, geophysical contours shaped by rivers, valleys and mountains, the historical trajectory, demography and social mobility, and also economic aspects of speech. These are all great influences on the spread of differences in language communities, mainly lexical, and thus they banish an alleged linguistic unity to the arid terrain of the educational policies established by each government and at other levels of language.

From the perspective of Portuguese-speaking communities across the globe, the lexical choices, instead of unifying, in fact strongly characterise each linguistic variety, serving as a shopfront for its very idiosyncrasies. Some may even serve as lexicultural stereotypes, such as *carnaval*, *bacalhau*, *baobá* or *matabicho*.

Many words of the Portuguese language in Portugal, such as *agrafador, boleia*, *autocarro*, *fita-cola, biberão, bica, tira-cápsulas, mulher-a-dias, telemóvel, melga, penso, penso-rápido, fato, miúdo* and *guarda-redes* find their equivalents in Brazil with *grampeador, carona, ônibus, mamadeira, café, saca-rolhas, diarista, celular, muriçoca, absorvente, bandeide, paletó, menino* and *goleiro*.

In regard to Portuguese in Africa, Bacelar do Nascimento (2006) identified lexical variants of Portuguese varieties from Angola, Cape Verde, Guinea-Bissau, Mozambique and São Tomé e Príncipe, based on a combined corpus of some 3,200,124 tokens from the written and oral modes. While the weight of these two modes is disproportionate, oral data representing only four percent of the data, his study reveals that the scale of lemma recurrence in the observed varieties is less common than one might suppose, since only 25 % of the total number of lemmas in the corpus are present in all five varieties; further 10 % are found in at least four varieties; another 11 % in at least three; further 12 % in two, leaving 38 % of the lemmas which are not shared between any of the varieties (Bacelar do Nascimento 2006, 195). This indicates a significant percentage of lack of lexical unit in this continent, too.

---

45 "A esse entrecruzar-se de dialetos sociais, espaciais e de normas linguísticas impõem-se as normas específicas da língua escrita que neutralizam muitas das diferenças da fala quotidiana, mas estão longe de anulá-las. A espinha dorsal que, entretanto, une todas essas diferenças se capta em um nível de abstração maior, que é do sistema de regras comuns que subjaz a essas diferenças, e que dá suporte a que, enquanto fenômeno histórico se possa afirmar que nesses diferentes pontos do globo está, ali, a língua portuguesa e não outra língua".

Nowadays, ongoing research is currently addressing what has long been an urgent need, that is, the need for more accurate knowledge of the lexicon in use in Portuguese, in the dimension of its varieties. Under the responsibility of the International Portuguese Language Institute (IILP) and coordination of the CPLP (Portuguese Speaking Countries Community), the development of the Common Orthographic Vocabulary of the Portuguese Language (VOC), as provided by the Spelling Agreement established, currently combines 250,000 (two hundred and fifty thousand) entries, with more than 2,000,000 (two million) spellings. It intends to gradually integrate the national spelling vocabularies of each country member within a common methodology. To date, only Brazil, Portugal and Mozambique have added their vocabularies to the new database.

Associated with this initiative of the VOC is the strategy of promotion, distribution and projection of the Portuguese language, as proposed by the Brasília Action Plan (*Plano de Ação de Brasília*) of 2010, which is based on the following main points: i) implementation of the Portuguese language in international organisations; ii) promotion and dissemination of Portuguese language teaching in the spaces of the CPLP, in the spaces of the diaspora, as well as in teaching Portuguese as a foreign language (L2); iii) monitoring the implementation of the spelling agreement; iv) public dissemination of the Portuguese language with regard to cooperation agreements, publishing, cultural activities, exchanges, among other actions; v) inclusion of civil society in all the proposed aspects.

If these goals are properly fulfilled, the adoption of such measures may allow greater integration of communities and the empowerment of the Portuguese language in relation to other more widely spoken languages around the world, providing that the different cultures that currently use it become seen together in the world, beyond what has been marked by the lexicon in their literature:

> TRUTH, LIES, certainty, uncertainty...
> That blind man over there on the road knows these words too.
> I'm sitting on the top step and I have my hands clasped
> On the higher of my crossed knees.
> Are truth, lies, certainty, uncertainty the same?
> Something changed in part of reality—my knees and my hands.
> What science has knowledge for this?
> The blind man goes on his way and I don't make any more gestures.
> It's not the same time any more, or the same people, nothing is the same.
> This is being real (Pessoa 1977, 232).[46]

46 "VERDADE, MENTIRA, certeza, incerteza... / Aquele cego ali na estrada também conhece estas palavras. / Estou sentado num degrau alto e tenho as mãos apertadas / Sobre o mais alto dos joelhos cruzados. / Bem: verdade, mentira, certeza, incerteza são as mesmas? / Qualquer coisa mudou numa parte da realidade—os meus joelhos e as minhas mãos. / Qual é a ciência que tem conhecimento para isto? /

To paraphrase Pessoa, one of the greatest poets of the Portuguese language, one must say that to history, what is real is change; truth, lies, certainties and uncertainties may one day no longer be, but there is the lexicon for knowing a word.

6 References

Alkmim, Tânia/Petter, Margarida (2008), *Palavras da África no Brasil de ontem e de hoje*, in: José Luiz Fiorin/Margarina Petter (edd.), *África no Brasil: a formação da língua portuguesa,* São Paulo, Contexto, 135–177.

Bacelar do Nascimento, Maria Fernanda (2006), *Corpora comparáveis e variação lexical nas variedades africanas do português,* Alfa 50, 189–204.

Barbosa, Maria Aparecida (2000), *Dos processos de engendramento e manifestação do neologismo nos discursos essencialmente figurativos*, in: José Carlos de Azeredo, *Língua portuguesa em debate: conhecimento e ensino*, Petrópolis, Vozes, 176–191.

Barcellos da Silva, Maria Emília (2000), *O dinamismo lexical: o dizer nosso de cada dia*, in: José Carlos de Azeredo (ed.), *Língua portuguesa em debate: conhecimento e ensino*, Petrópolis, Vozes, 142–146.

Biderman, Maria Tereza ([2]2001), *Teoria linguística*, São Paulo, Martins Fontes.

Bonvini, Emilio (2002), *Palavras de origem africana no português do Brasil. Do empréstimo à integração*, in: José Nunes/Margarida Petter (edd.), *História do saber lexical e constituição de um léxico brasileiro*, São Paulo, Humanitas, 147–162.

Calvet, Louis-Jean (1974), *Linguistique et colonialisme: petit traité de glottophagie*, Paris, Payot.

Cardoso, Suzana Alice, et al. (2014a), *Atlas linguístico do Brasil*, vol. 1, Londrina, Eduel.

Cardoso, Suzana Alice, et al. (2014b) *Atlas linguístico do Brasil*, vol. 2, Londrina, Eduel.

Cardoso, Wilton/Cunha, Celso (1978), *Estilística e gramática histórica*, Rio de Janeiro, Tempo Brasileiro.

Coelho, Antônio Borges (2000), *Os argonautas portugueses e o seu velo de ouro: séculos XV e XVI*, in: José Tengarrinha (ed.), *História de Portugal*, Bauru/São Paulo, EDUSC/UNESP, 57–75.

Coutinho, Nilton Ribeiro (2010), *Dicionário, palavras oriundas de línguas que mais contribuíram para a formação da língua portuguesa falada no Brasil: africanas, árabe, espanhola, francesa, inglesa, italiana e tupi*, Salvador, Quarteto.

Cunha, Antônio Geraldo da (1978), *Dicionário histórico das palavras portuguesas de origem tupi*, São Paulo, Melhoramentos; Brasília, Universidade de Brasília.

Cunha, Celso (2004), *Sob a pele das palavras: dispersos*, Rio de Janeiro, Nova Fronteira/Academia Brasileira de Letras.

Drummond de Andrade, Carlos (1985), *Contos plausíveis*, Rio de Janeiro, José Olympio.

Drummond de Andrade, Carlos ([30]2005), *A rosa do povo*, Rio de Janeiro, Record.

Dubois, Jean, et al. (1994), *Dictionnaire de linguistique et des sciences du langage*, Paris, Larousse.

Fename/Ministério da Educação e Cultura (1956–1978), *Dicionário escolar da língua portugues*a, Rio de Janeiro, Fename.

Ferreira, Aurélio Buarque de Holanda (1999), *Dicionário Aurélio Eletrônico Século XXI*, versão 3.0, Rio de Janeiro, Nova Fronteira.

O cego continua o seu caminho e eu faço mais gestos. / Já não é a mesma hora, nem a mesma gente, nem nada igual. / Ser real é isto".

Garcez, Pedro/Zilles, Ana Maria (2002), *Estrangeirismos*, in: Carlos Alberto Faraco (ed.), *Estrangeirismos: guerra em torno da língua,* São Paulo, Parábola, 15–36.
Gonçalves, Perpétua/Chimbutane, Feliciano (2009), *Assimetrias da mudança linguística em situação de contacto de línguas: o caso do português e das línguas bantu em Moçambique*, in: Ana M. Carvalho (ed.), *Português em contacto*, Madrid/Frankfurt am Main, Iberoamericana/Vervuert, 31–51.
Guilbert, Louis (1973), *Théorie du néologisme*, Cahiers de l'Association internationale des études françaises 25, 9–29.
Guimarães Rosa, João (1967), *Hipotrélico*, in: João Guimarães Rosa, *Tutaméia: terceiras histórias*, Rio de Janeiro, Livraria Olympio, 64–67.
Houaiss, Antônio (2001), *Dicionário eletrônico Houaiss da língua portuguesa*, Versão 1.0, Rio de Janeiro, Objetiva.
Instituto Brasileiro de Geografia e Estatísica (2000), *Brasil: 500 anos de povoamento*, Rio de Janeiro, IBGE, 226.
Machado Filho, Américo Venâncio Lopes (2009), *Um flos sanctorum trecentista em português*, Brasília, Editora UnB.
Machado Filho, Américo Venâncio Lopes (2010), *Um ponto de interseção para a dialectologia e a lexicografia: a proposição de um dicionário dialetal brasileiro com base nos dados do AliB*, Estudos Linguísticos e Literários 41, 49–70.
Machado Filho, Américo Venâncio Lopes (2013), *Dicionário etimológico do português arcaico*, Salvador, Edufba.
Machado Filho, Américo Venâncio Lopes (2014), *Do conceito de "variante" nos estudos do léxico de perspectiva histórico-variacional*, Filologia e Linguística Portuguesa 16/2, 261–275.
Mattos e Silva, Rosa Virgínia (1988), *Diversidade e unidade: a aventura linguística do português*, Revista ICALP 12/13, 13–28.
Mattos e Silva, Rosa Virgínia (2013), *A tessitura do léxico de uma língua histórica*. in: Rosario Álvarez et al., *Ao sabor do texto: estudos dedicados a Ivo Castro*, Santiago de Compostela, Universidade de Santiago de Compostela, 403–419.
Medeiros, Carmen Lúcia de (2008), *Desfazendo as tranças: estudo renovado da interferência das línguas africanas no léxico do português brasileiro,* Undergraduate degree thesis, Salvador, Universidade Federal da Bahia.
Mendonça, Renato (1948), *A influência africana no português do Brasil*, Porto, Livraria Figueirinhas.
Moreira, Marco Antônio (2002), *A teoria dos campos conceituais de Vergnaud, o ensino de ciências e a pesquisa nesta área*, Investigações em Ensino de Ciências, 7, 7–29.
Mussa, Alberto (1991), *O papel das línguas africanas na história do português do Brasil*, Master's thesis, Rio de Janeiro, Universidade Federal do Rio de Janeiro.
Nascentes, Antenor (1951), *Dicionário etimológico da língua portuguesa*, Rio de Janeiro, Francisco Alves.
Neto, Serafim da Silva (1960), *A língua portuguesa no Brasil. Problemas*, Rio de Janeiro, Liv. Acadêmica.
Pessoa, Fernando (1977), *Obra poética*, Rio de Janeiro, Nova Aguilar.
Pessoa, Fernando (1980), *Textos de crítica e de intervenção*, Lisboa, Ática.
Pessoa De Castro, Yeda ([2]2005), *Falares africanos na Bahia: um vocabulário afro-brasileiro*, Rio de Janeiro, Academia Brasileira de Letras.
Piel, Joseph-Maria (1989, [1]1976), *Origens e estruturação histórica do léxico português*, in: *Estudos de Linguística Histórica Galego-Portuguesa*, Lisboa, IN-CM, 9–16.
Pottier, Bernard (1974), *Linguistique générale: théorie et description*, Paris, Klincksieck.
Ramos, Jânia/Venâncio, Renato (2002), *Topônimos mineiros: uma fonte para a história social da língua portuguesa*, in: Eugênia Duarte/Dinah Callou, *Para a história do português brasileiro*, vol. IV, Rio de Janeiro, UFRJ/Letras, 113–123.

Ribeiro, Darcy (1997), *O povo brasileiro. A formação e o sentido do Brasil*, São Paulo, Companhia das Letras.
Schneider, John (1991), *Dictionary of African borrowings in Brazilian Portuguese*, Hamburg, Buske.
Soares, Antônio Joaquim de Macedo (1954), *Dicionário brasileiro da língua portuguesa*, Rio de Janeiro, Instituto Nacional do Livro.
Tinhorão, José Ramos (1997), *Os negros em Portugal. Uma presença silenciosa*, Lisboa, Caminho.
Viaro, Mário (2004), *Por trás das palavras. Manual de etimologia do português*, São Paulo, Globo.
Viaro, Mário (2011), *Etimologia*, São Paulo, Contexto.
Viaro, Mário, et al. (2015), *Manual do NEHiLP*, https://www.usp.br/nehilp/infos/manual.pdf, last accessed 16.02.2022.

Charlotte Galves and Marilza de Oliveira

8 The debate on Brazilian and European Portuguese

Abstract: In this chapter, we focus on the relationship between Brazilian and European Portuguese, reviewing and discussing the arguments and facts that have been presented as bearing on the issue. We start out by establishing the main linguistic differences between the two varieties, discussing the geographical origin of the various aspects that divide them. Then, the chapter provides an overview of the history of the debate on the nature of Brazilian Portuguese, during the 19th and 20th centuries. Finally, we discuss the topic from a theoretical point of view of contemporary language, presenting arguments that support the claim that Modern European and Brazilian Portuguese correspond to two different linguistic competences.

Keywords: dialectology, linguistic identity, I-language/E-language, purism, nationalism

1 Introduction

Today, Brazil is the largest Portuguese-speaking country. In spite of the fact that Tupi-based lingua francas were spoken for centuries, and that the language of the European colonists was acquired and transmitted by millions of African slaves under conditions that may have favored the emergence of Creole languages, the language brought by the Caravelas eventually spread throughout the country. Moreover, one can argue that it has been sufficiently preserved to the point that it can legitimately keep the same name five centuries later. Differences, however, soon became apparent, and were acknowledged as soon as the 18th century. After the country became independent in 1822, the question of the integrity/specificity of the Brazilian way of speaking the language of the former colonizers emerged as a social, educational and political issue.

Linguistic historiography has adopted the Saussurean dichotomy *langue* and *parole* as well as the Coserian trichotomy *system*, *norm* and *speech* in order to allocate Brazilian Portuguese in the supposedly unitary and homogeneous linguistic space of Portuguese, giving the Brazilian variety the status of *speech* or the status of a *norm* within a homogeneous system (Vázquez Cuesta/Mendes da Luz [2]1971, I, 129). However, it is necessary to keep in mind that every act of speech—every concrete linguistic activity—reflects both the system and the norm, abstractions that are distinguished respectively by the functional character and the social and cultural character of linguistic elements. Within the proposals of these eminent linguists, *speech*

https://doi.org/10.1515/9783110405958-009

is neither a deviation nor an autonomous or subaltern reality of the abstract levels. Therefore, it cannot be taken as a guideline to determine positions in linguistic hierarchical schemes.

As for the abstract levels, although one single system may contain several norms, it is not possible to limit the difference between European Portuguese (henceforth EP) and Brazilian Portuguese (henceforth, BP) to the level of the *norm* (in Coseriu's sense). It is not that just social and cultural factors separate the two linguistic varieties. The difference is more profound, and it is also systemic, as Kabatek (2015) points out.

In fact, the differences between European Portuguese (henceforth EP) and Brazilian Portuguese can be observed on all linguistic levels. Following Révah's (1956) discussion concerning the phonetic phenomena that distinguish the two varieties, it is important to emphasize that such differences may have different origins. Taking as the point of departure BP and EP's common ancestor, Classical Portuguese (henceforth ClP), it is sometimes the case that both varieties underwent changes in different directions. Other times, only one of them innovated and the other kept the old form or structure. In this case, BP is frequently the innovator, but not always. We shall try to make this aspect clear whenever possible, since this question has been the subject of many claims and debates, not always founded on objective facts. Here, we shall concentrate on morpho-syntactic differences, addressing more precisely the question of the innovation versus conservatism on this level. Concerning differences in pronunciation and the lexicon, we refer the interested reader, respectively, to ↗4 Historical phonetics and phonology, and to ↗7 The history of the lexicon.[1]

It is important to remember in this context that the status of a linguistic variety does not derive directly from its structural particularities, but rather from ideological issues connected to linguistic and educational policies. These establish a linguistic matrix around which the other varieties orbit forming a linguistic unit. The status also depends on the feeling of belonging as well as the intellectual will of linguistic communities to participate in this game.

One of the ideological issues that underpinned linguistic policies in the 19th century was based on models of cultural analysis, a scenario in which the treatment given to the linguistic facts that typified BP made references to miscegenation. Silvio Romero's (1888) idea that the Brazilian cultural sovereignty depended on racial hybridization was revived in favor of the idea of a Brazilian dialect. From another point of view, although not following an orchestrated plan, the paradigm of social evolutionism, with European civilization in its corner (Elias 1990), had linguistic purism on the agenda.

Amid the linguistic turmoil, there was the colonial Portuguese language and Standard Portuguese in Brazil, whose classic profile was condemned by Portuguese philol-

1 Cf. also Noll (2008, 7.1, 3.3), and specifically regarding the phonetics/phonology, Révah (1956), Teyssier (1980), Frota/Vigário (2001), and Sandalo et al. (2006).

ogists and grammarians who, accustomed to the "curve of civilization", intended to update it according to Lusitanian models. Social evolutionism removed Europe from the civilized corner and the surge of the nationalist paradigm (Hobsbawn 2008, [1]1990) created the conditions for developing the Brazilian identity. In this situation, the adjective "national" was used for a long period to designate the language spoken in Brazil, appeasing defenders of the Portuguese language and creating conditions for recognizing a standard Brazilian variety within the Portuguese domain.

These cultural changes underlie the studies of specific linguistic characteristics and label proposals for the language used in Brazil, having a range of names throughout the referential process: Brazilian language, Brazilian dialect, vernacular language, Brazilian Style, national idiom and national language. In this chapter, we intend to build upon studies that have addressed the nature of the language spoken in Brazil within the Portuguese-Brazilian range of reference and unveil the web of meanings woven by philologists/grammarians, and to which they are tied (Geertz 1989).

The remainder of this chapter is organized in 4 sections. Section 2 briefly describes the main morpho-syntactic differences between European Portuguese and Brazilian Portuguese. The debate on the identity of Brazilian Portuguese is the object of Section 3. In Section 4, we propose a new way of addressing the issue of the difference between the two varieties. Section 5 concludes the chapter.

2 The main differences between European Portuguese and Brazilian Portuguese

Much has been written on the morpho-syntactic differences between BP and EP, as we shall see throughout this section. The most salient ones have been, in general, the focus of normative considerations, having been abundantly described and analyzed from a linguistic point of view.[2] Others have only recently been acknowledged and discussed. In this section, we shall first consider the ones that have been emblematic of the difference over time: clitic placement, agreement, and subject-verb order. We shall focus, when pertinent, on the normative debates or the discussions of their origins. We shall then review other phenomena that, although intimately related to the aforementioned ones, were not taken into special consideration until the emergence of linguistic studies. This is the case of null arguments, the morpho-syntax of personal pronouns, and the use of definite determiners. We shall conclude with a summary of the innovations in each case.

2 For a systematic comparative approach, cf., among others, Teyssier (1976), Galves (1987; 1998), Kato/Peres (2005), Vieira/Brandão (2014), and Dubert/Galves (2016).

For more details, we refer the reader to ↗5 Historical morphology, ↗6 Historical syntax, ↗13 Morphology and ↗14 Syntax.

2.1 Clitic placement

Clitic placement deserves to be addressed first on this list, since it played a central role in the early discussions on Brazilian linguistic identity. The issue was so important that, at the very beginning of the 20th century, the review of the Chamber of Deputies' Civil Code project by the famous lawyer Ruy Barbosa mainly concerned the grammatical correctness of the text, with this issue being at the core of reviewers' criticisms (cf. Pagotto 2011). Barbosa's comments initiated a debate that lasted years (Barbosa 1949, [1]1902; 1969, [1]1902).

The issue of clitic placement is in large part due to a change that occurred in European Portuguese from the 18th century on: in the contexts in which clitic pronouns could appear either before the verb (proclisis) or after the verb (enclisis) in ClP, enclisis became obligatory. In historical records, this change was complete by the end of the 19th century (cf. Martins 1994; Galves/Britto/Paixão de Sousa 2005). In BP, however, proclisis continued to be the normal placement, and this created a very salient difference between the two varieties. Since EP was considered the norm, educated Brazilians tended to copy it, against their natural tendency toward proclitic placement, and even exaggerated the use of enclisis. Only from 1920 on did proclisis increase in texts (cf. ↗6 Historical syntax). This situation was complicated by the fact that BP innovated by extending the proclitic pattern to the absolute first position of clauses, in opposition to the strict prohibition of clitic pronouns in first position (the so-called Tobler Mussafia Law), which had been active in Portuguese throughout European history. The example below shows a revision of Paulo Coelho's text in the Portuguese edition of *O Alquimista* 'The Alchemist' (cf. Galves/Ribeiro/Torres Moraes 2005).

(1) a. Me chamo Fátima – disse a moça... (BP)
cl1 call.1sg Fatima said the girl
b. Chamo-**me** Fátima – disse a moça (EP)
call .1sg–cl1 Fatima said the girl
'My name is Fatima, said the girl'

We must emphasize, however, that proclitic placement is not the only difference between EP and BP. They also differ with respect to the category to which the clitic attaches. In the former, it adjoins to the tensed form, be it an auxiliary or the thematic verb. In the latter, it is proclitic to the thematic verb, be it tensed or not. This is exemplified by another revision made in the Portuguese edition of Paulo Coelho's novel *O Alquimista*.

(2) a. Como tinha **se** comportado de maneira correta... (BP)
as had CL3.REFL behaved in manner correct

b. Como **se** tinha comportado de maneira correcta...(EP)
as CL3.REFL had behaved in manner correct
'As she/he had behaved correctly...'

Summing up, we can say that the split between EP and BP regarding the syntax of clitics derives from innovations on both sides. But it is interesting to note that the more observable difference—enclisis vs. proclisis—may not be the most important one from a deeper grammatical point of view. We shall return to this point in Section 2.5, when we consider the syntax of pronouns as a whole.

2.2 Verbal and nominal agreement

The partial loss of verbal and nominal morphological agreement, respectively illustrated in (3) and (4), has been one of the more frequently studied (cf. Guy 1981; Scherre 1988; Holm 2004; among others) and stigmatized features of BP.

(3) eles bebe, fica cantando
they drink.3SG keep.3SG singing
'they drink and sing' (Naro/Scherre 2007, 72)

(4) um dos mais velho orixás
one of-the.PL more old.SG deities
'one of the more ancient deities' (Holm 2004, 101)

At the end of the 20th century, verbal agreement occurred at a rate ranging from 16 % in isolated Afro-Brazilian communities to 73 % among educated people from the city of Rio de Janeiro (cf. Lucchesi/Baxter/Ribeiro 2009, 348). This suggests that the morphological realization of agreement varies along three axes: isolated vs. integrated populations, non-urban vs. urban contexts, less educated vs. more educated. From the very beginning of studies on "overseas" dialects of Portuguese (cf. Coelho 1967, [1]1883) to the present day, this issue has been a topic of discussion and a source of debate. It led some researchers to argue that BP underwent a creolization phase (Guy 1981), the effect of contact with indigenous and African languages. Such a claim has been much debated, cf. ↗3 The history of linguistic contact underlying Brazilian Por-

tuguese.[3] Some years ago, Naro/Scherre (2007) argued that the tendency to lose morphological agreement markers is rooted in European Portuguese. A decisive piece of evidence, according to the authors, is that, despite differences in frequency, the contexts that favor the absence of agreement are the same in EP and BP. Based on a comparative study on Portuguese and Brazilian urban dialects, Vieira/Brandão (2014) counter that such contexts are universal. They further argue that the quantitative profile of the variation in agreement is decisively different in each variety.[4] In Section 2.3, we present agreement facts specific to BP.

As far as person features are concerned, it is important to note that in most dialects in Brazil the second person singular is no longer used in the verbal paradigm. This creates a context for the lack of subject-verb agreement for speakers who use the 2nd person pronoun *tu* instead of, or in variation with, the 3rd person pronoun *você*, which in Brazil has lost the polite/formal feature, while in Portugal, it remains (cf. ↗6 Historical syntax). We return to the grammatical expression of the 2nd and 3rd person in Section 2.5.

2.3 The syntax of subjects

Both EP and BP underwent a change with respect to the position of subjects. In ClP, the VS(O) order was very frequent, due to the movement of the verb to a high position, arguably Comp, in this language (cf. Antonelli 2011). Furthermore, in this position, the subject was not obligatorily interpreted as focus, i.e. new information, as it is nowadays (Galves/Gibrail 2018). In EP, an abrupt change appeared in texts by the first generation of 18th-century authors (Paixão de Sousa 2004; Galves/Paixão de Sousa 2017). From then on, VSO became marginal, and VS was restricted to unaccusative verbs. In EP, VSO is only felicitous in contexts in which both the subject and object were new information (Costa 2004). BP went further in the same direction. VSO fell into disuse and SV was preferred even with unaccusative verbs (cf. ↗6 Historical syntax for a survey of diachronic data).[5] It also became the canonical order in interrogative sentences.

Moreover, BP has new pre-verbal subjects. These were first described by Pontes (1987). She showed that BP shares with "topic-oriented" languages constructions in which the first constituent DP of the sentence is not the semantic subject of the verb, but some locative or genitive argument, immediately followed by the verb and the internal argument of the verb:

---

3 Cf. also the discussion in Silva Neto ([1]1950), Tarallo (1993a; 1993b), Pagotto (2007), Lucchesi/Baxter/Ribeiro (2009).

4 For a discussion of Naro/Scherre (2007), cf. also Galves (2012) and Lucchesi (2012).

5 The inverted order, however, maintained a prestigious status in official texts. In the minutes of the *Normal School*, written between 1900 and 1920, Ribeiro (2015) finds, respectively, 45 % and 22.6 % postposed subjects with mono-argumental and bi-argumental verbs.

(5) Esta casa bate muito sol
this house strikes much sun
'There is too much sun on this house'

(6) A Sarinha nasceu o dente
the Sarinha (was) born the tooth
'Sarinha's tooth came in.'

As discussed recently by many scholars (Negrão/Viotti 2008; Avelar/Galves 2011; Munhoz/Naves 2012; Andrade/Galves 2014; among others), there can be overt agreement between this pre-verbal NP and the verb:

(7) Estas casas batem demasiado sol
these houses strike.3PL too-much sun
'There is too much sun on these houses'

This type of agreement is absolutely impossible in EP (Costa 2010), and therefore, could not be attributed to a Portuguese drift. Recent studies have shown that the same phenomenon exists in Bantu languages. This suggests that this kind of sentence could be the effect of contact between Portuguese and those languages in Brazil (cf. Avelar/Galves 2014).

2.4 Null subjects

Diachronic studies on null subjects indicate that EP and BP evolved differently with respect to this property. While in the former, no significant difference seems to have emerged over time, Tarallo's (1983) pioneering work showed that the use of null subjects was gradually decreasing in the latter. Furthermore, subsequent work showed that qualitative differences accompanied the quantitative ones. One is the possibility of interpreting null subjects of tensed clauses as indeterminate, as shown by the contrast between (8a) and (8b):

(8) a. Aqui conserta sapatos (BP)
here repair.3SG shoes
b. Aqui conserta(m)-se sapatos (EP)
here repair.3SG(PL)-REFL shoes
'Here shoes are repaired.'

The facts illustrated in (8) raise the question of the causal relationship between a and b, which can be expressed in the following terms: Is the interpretation of (8a) as inde-

terminate due to the general tendency for the medio- reflexive clitic pronoun SE to not be pronounced (cf. Galves 1987; Nunes 1990) or is the loss of SE due to the possibility of interpreting a. as indeterminate? We know that SE also tends to be dropped in its other uses, particularly in pronominal verbs, as in *Ele sentou (-se)* 'He sat down' and in causative alternation verbs, as in *Eu abri a porta*/*A porta abriu (-se)* 'I opened the door/the door opened'. However, it continues to be used in reflexive contexts. We can therefore think that the disuse of SE to express indeterminacy is correlated with the new syntax of null subjects, which in turn derives from the peculiarity of the third person in BP. Recent work on null subject languages has shown that other languages display the same set of null subject properties as BP (cf. Holmberg/Nayudu/Sheehan 2009). Such languages have been dubbed Partial Null Subject Languages. We return to this question in Section 4.

2.5 The morpho-syntax of pronouns

The BP pronominal paradigm and syntax present several characteristics that constitute strong innovations not only in Portuguese history, but in the Romance domain as a whole. First and second person accusative and dative clitic *me*/*te* 'me'/'you' pronouns vary with strong pronouns, as exemplified in (9). The third person accusative clitic *o*/*a* 'him'/'her' is substituted in speech by the strong nominative form *ele*, or by an empty category, as illustrated in (10). In EP, as in other Romance languages, the use of strong pronouns in object position requires the presence of the preposition *a* 'to', and clitic doubling of the pronoun (*Vi-**o a** ele*). Furthermore, in EP, this construction entails a contrastive focus interpretation, which is not the case in BP.

(9) te **vi** / vi **você** (ontem)
CL2SG saw / saw you yesterday
'I saw you yesterday'

(10) (Não) **o** vi / vi **ele** / vi **ec** [ec=empty category]
NEG CL3SG saw / saw him / saw ec
'I did not see him'

Recent work has shown that the variation in (9) is stable,[6] in the sense that the innovation has not led to the loss of the old form, since the clitic still continues to be frequently used (cf. Souza 2014). This strongly contrasts with what has happened with

6 It is worth noting that the variation between clitic and tonic pronouns is more dialectally constrained in the 1st person than in the 2nd person.

the 3rd person, where clitics are no longer used. In their place, we find tonic pronouns, and, more frequently, null objects (Duarte 1986). We shall see in Section 4 that EP also has null objects, although in a more restricted way (Raposo 1986). In both varieties, this is an innovation, since null objects are not found in ClP. However, while null objects vary with 3rd person clitic pronouns in EP, the former have replaced the latter in BP.

Lastly, BP differs from EP in that the third person dative clitic *lhe* tends to be used only as a form of address, corresponding to the strong pronoun *você* 'you', meaning exclusively 'to you'. As a consequence, *lhe* and *te* freely vary (ex. 11). Accordingly, the possessive pronouns *teu* and *seu* freely vary to refer to *você* (ex. 12).

(11) Eu **te/lhe** dei esse livro
I CL2/CL3Dat gave this book
'I gave you this book.'

(12) Você trouxe **teu/seu** livro?
you brought poss2/poss3 book
'Did you bring your book?'

This can be taken as a consequence of the complete loss of any distinction between the use of *tu* and *você*, in contrast with what happens in EP, and the replacement of the former by the latter in most dialects (cf. ↗6 Historical syntax, and references therein). Note that this loss is grammatical as well as discursive, since in most dialects, the 2nd person pronoun *tu* agrees with the 3rd person of the verb. This brings us back to the issue of the "loss" of agreement, and the associated question of the creolization process. An alternative hypothesis would be that this is the effect of the influence of Bantu languages, such as Kimbundu, which do not morphologically differentiate the 2nd and 3rd persons.[7]

2.6 The use of determiners

Another interesting difference between BP and EP concerns the use of definite determiners. Portuguese as a whole differs from other Romance languages in licensing noun phrases without determiners (*Bare nouns*) in all syntactic contexts, particularly in subject position (see Kabatek/Wall 2013 and references therein). This is frequently

7 The importance of Kimbundu during the early centuries of the colonial period in Brazil is emphasized by many scholars (cf. Bonvini 2008), and is well illustrated by the fact that the first grammar of the language was written in Bahia at the end of the 17th century by the Jesuit priest Pedro Dias.

found in ClP. In the evolution from ClP to EP, we observe a tendency to use definite determiners more systematically. The opposite occurred in the history of BP, in which, in subject position, plural bare nouns can receive generic as well as existential interpretations, and singular bare nouns are possible in subject and object position with a generic interpretation.[8] In EP, the same uses exist but in a much more restricted way, and generic NPs tend to be expressed with a definite article (cf. Peres 2013). The tendency to use more definite articles is also observed in the case of possessive NPs, where this use is variable in BP, as in ClP, but is categorical in EP.

2.7 Summarizing

The differences between EP and BP derive from both Portuguese and Brazilian innovations. As for the former, we mentioned two of them:[9] obligatory enclisis in the aforementioned proclisis/enclisis variation contexts (cf. 2.3) and the obligatory use of a determiner with possessive pronouns (cf. 2.6). A third one worth mentioning is the replacement of the gerundive periphrasis to express progressive aspect with the periphrasis *a* + infinitive, as illustrated in (13) from Teyssier (1976).

(13) a. Está se fazendo tarde
be.PRS CL.REFL do.GER late

b. Está-se a fazer tarde
be.PRS CL.REFL to do.INF late
'It is getting late'

In the Tycho Brahe Corpus (Galves/Andrade/Faria 2017), the first examples of the b construction appear in authors born in the 18th century. The two forms first co-exist, but the new one eventually wins the competition. From a formal point of view, regarding this construction, BP is therefore the conservative variety. However, if we consider usage, we observe innovations as well, since the gerundive periphrasis in BP has been replacing simple tenses in cases in which no progressive aspect is involved.

There is no doubt that there are more Brazilian innovations than Portuguese innovations. They affect clitic-placement, the morpho-syntax of pronouns, agreement facts and the interpretation of null subjects. As for the phenomena resulting from innovations in both varieties, both when they converge and when they diverge, we observe that BP went further in differentiating from the original syntax than EP did. We

8 Empirical evidence for those claims can be found in Schmitt/Galves (2016).

9 Another important change in EP affected its rhythmic pattern (cf. references in fn. 2). Galves/Paixão de Sousa (2017) argue that this change was at the origin of the generalization of the SV-cl pattern and the generalization of SV with non-unaccusative verbs.

saw that this was the case for null objects, which substitute 3rd person clitic pronouns, while in EP, they coexist, and for SV order, which affects even unaccusative verbs, and extends to the emergence of topic subjects. This fact can be expressed in terms of the occurrence of more parametric changes in BP (cf. Galves/Kroch 2016).

3 The history of the debate

Studies on grammar, at the end of the eighteen-hundreds, underwent abrupt upheavals due to the change in models for linguistic analysis: the logical and speculative model gave way to the scientific and historical model, which, at the time, was dominated by the perspective of historical-comparative and classification investigations[10]. The linguistic discussions resulting from this perspective reached a program for teaching Portuguese proposed by Fausto Barreto[11] and the production of school grammar books. They were innovative for including Brazilianisms, but they were grouped into a table of errors and bad language habits.

This interpretation followed the paradigm of social evolutionism in its different aspects. In this context, the treatment of the linguistic facts that were characteristic of Portuguese in Brazil under the label "Brazilian dialect" made references to miscegenation. On the other hand, the same model of cultural analysis, with European civilization at the vertex (Elias 1990), explained the waves of linguistic purism.

Amid this cultural turbulence was educated Brazilian Portuguese, whose classical profile was condemned by Portuguese philologists and grammarians (henceforth, intellectuals) who, accustomed to the "curve of civilization", wished to modernize it according to Lusitanian models. With the end of social evolutionism, which removed Europe from the civilizing vertex, and with the start of the paradigm of nationalism,[12] which created the conditions for developing the notion of national identity, the expression "Brazilian dialect" gave way to "national language/idiom" to refer to the Portuguese language. Then, it changed to the designation "Portuguese language of/in Brazil", calming the waters between Portuguese and Brazilian intellectuals. Paradoxically, in the period marked by national identity, the designations "national language" and "Portuguese language", without further qualification, served those who rejected, as well as those who anticipated, recognition of a Brazilian linguistic variety.

These cultural changes support descriptive studies of specific linguistic characteristics and proposals on designating spoken Portuguese in Brazil, the history of which follows the path outlined by the formation of the Brazilian nation.

---

10 Altman (2004) analyzes the impact that this change had on the disciplines of Philology and Linguistics, in the period from the end of the 19th century until the mid-20th century.

11 This program is referenced by Castilho (s/d), Coelho/Silva (2018), Faraco (2018), among others.

12 The process of historical invention of the modern nations began with the equation nation-State-people stipulated in the French Declaration of the Rights of Man of 1795 (Hobsbawn 1990, [2]2008).

3.1 The linguistic issue in the discourse of inventing the Brazilian nation

Mapping the word "nation" in the political vocabulary from the 1830s on, Chauí (2000) provides the following time frame for the historical process of the invention of the Brazilian nation: i. "principle of nationality": this discourse links nation and territory (1830–1880); ii. "national idea" is linked to language, religion and race (1880–1918); iii. "national issue" underscores the national consciousness (1918–1960). The first two periods were governed by the idea of the "national character"; the last was supported by the idea of a "national identity".

Adapting the time frame of the invention of the Brazilian nation to the linguistic issue, we will assume the existence of two periods: i. national character and ii. national identity. The discourse on the Brazilian national character includes two phases. The first covers the period of the formation of the Empire and of the linguistic controversies involving Brazilian intellectuals. It is characterized by the perception of linguistic features specific to Brazilian speech, called "Brazilianisms", and by an appreciation of the indigenous peoples. In this phase, Afro-descendants were ignored (Fiorin 2016).

The second phase of building the discourse on the Brazilian character includes Afro-descendants, but from the perspective of social evolutionism and, in particular, criminal anthropology (Lombroso [3]1884, [1]1876), equating miscegenation with criminality. In this context, Brazilianisms started to be interpreted as errors or remnants of an interrupted creolization process, and the Brazilian dialect designation took on a connotation of ignorance.

In the period of developing the national identity, the dialect label was reduced to regional speech and substituted by a national language that abstracted linguistic differences[13]. Paradoxically, the emergence of a national consciousness occurred in parallel to the emphasis of a Luso-Brazilian brotherhood expressed by the Lusotropicalism of Freyre (1933), and the differences were mitigated by a view of a multicultural and multicontinental community. This (re)invention of a cultural subconscious communion between two populations created the conditions for substituting the label "national language" with "Portuguese of/in Brazil".

3.2 In the beginning, there were "Brazilianisms"

Nationalist movements loomed over Brazil despite its singular emancipation, on the occasion of the arrival of D. João VI, who elevated it to the status of United Kingdom.

13 In this period, there is a large number of papers entitled simply *Gramática Portuguesa* ("Portuguese Grammar").

These movements were also mitigated on the occasion of its independence (1822). The establishment of a monarchy governed by the heir of the Lusitanian Crown made the creation of the National State smoother. The European cultural atmosphere was imported not according to nationalist models, but in the form of recycling the elite, "modern consumers fascinated by imported luxury". The result was a delay in the perception of nationality, since 'between the institutionalization of the Nation State and the establishment of the nation as a territory and people, almost one century elapsed'[14] (Lessa 2008, 238).

The dynastic continuity of the House of Bragança corresponded to the preservation of the Portuguese social structure, heritage and linguistics, strengthened by the staying and continued entry of Portuguese and by the rise of some Brazilians on the social ladder. Even before independence, the presence of the Portuguese elite broadened the circle of the more well-endowed social class and made it possible to raise the consciousness of the Brazilian elite which, from the perspective of Silva Neto, began to assume the Court of Lisbon as a model.[15]

Moving the Court to Rio de Janeiro, the leveling of Brazil and Portugal (United Kingdom), the formation of the Empire, and the dynastic continuity created an environment conducive to waves of purism, which exercised 'a control more severe and fearsome than that of the Inquisition itself'[16] (Silva Neto 1977, 65), thus establishing a linguistic tyranny that legitimized only Lusitanian usage in the tropics.

The effervescent linguistic purism contrasted with the trend towards "manufacturing" a *Brazilian language*, a term which had already appeared in a decree in 1832.[17] In France, the Brazilian diplomat Domingos Borges de Barros (Visconde de Pedra Branca), between 1824 and 1825, indicated the existence of particular Brazilian linguistic characteristics[18] in an agenda that pursued 'not only new laws, but also a new consciousness'[19] (RAFDR 1922/182, 91, apud Schwarcz 1993, 141).

Brazilian nationalism was part of the elitist project of emancipation in cultural terms. In this program of inventing a model for the nation, the "Indianist" trend that had reached titles of nobility (Binzer 1994) and had provoked a change in Portuguese surnames in order to give them an exotic luster (Schwarcz 1998), diffused the linguistic issue. Literature (as well as fine arts) ennobled the indigenous, but denaturalized their voice. This critique of the poem by Gonçalvez de Magalhães ([3]1864) and the lin-

14 "entre a institucionalização do Estado Nacional e o delineamento da nação como território e povo, transcorreu-se quase um século".

15 An alternative view is defended by Oliveira (2019; 2021).

16 "uma fiscalização mais severa e temível que a da própria Inquisição".

17 The decree mentioned by Macedo Soares, *Dicionário brasileiro da língua portuguesa* 'Brazilian dictionary of the Portuguese language', apud Silva Neto (1950, 65).

18 In a paper published in *Introduction à l'Atlas etnographique du globe*, by Adrien Balbi (1826), apud Ribeiro (1979, 28–30).

19 "não apenas novas leis, mas também uma nova consciência".

guistic choices that José de Alencar made in *Iracema* (2006, [1]1865) triggered a major controversy. In his arguments, José de Alencar presented linguistic features of Portuguese in Brazil and used the nomenclature Brazilian dialect,[20] alternating with Brazilian language.

Among the critics of Alencar's linguistic expression, Pinheiro Chagas stood out, by condemning 'the lack of correction in the Portuguese language, or the obsession with turning Brazilian into a language different from Old Portuguese'[21] (apud Proença 1965, 198). Fierce critics of Alencarian language turned the exotic into error and moved the issue of language in Brazil to the forefront. Referred to as a Brazilian language as well as a Brazilian dialect by José de Alencar, tensions were ignited between Brazilian and Portuguese intellectuals regarding its designation.

The purist reaction occurred together with the founding of the philological society of Rio de Janeiro (1876) 'with the goal of fixing the language or making it return to the 16th century'[22] and the search for teachers in Lisbon for Brazilian schools (Silva Neto [1]1950, 212).

3.3 National idea: dialect, no!

The discussion on the linguistic object and its nomenclature was joined by other Brazilians. Gonçalves Dias (1921) criticized the purist idea that the vernacular nature of the Portuguese language was "intangible and sovereign". Counter to those who defended Lusitanian Portuguese as a model, João Ribeiro (1933, [1]1887) upheld the Portuguese language in the Brazilian case, with the exception that:

> 'Our grammar cannot be entirely the same as that of the Portuguese. Regional differences call for diverse styles and methods. The truth is that, by correcting them, we are in fact mutilating ideas and sentiments that are personal to us. It is no longer the language that we examine, it is our spirit that we subject to inexplicable servility. To speak differently does not mean to speak incorrectly'[23] (1933, [1]1887, 59).

20 Jeronymo Contador de Argote (1725) had already placed Brazilian among the overseas dialects (Sanches 1940, 152).

21 "a falta de correção na linguagem portuguesa, ou antes a mania de tornar o brasileiro uma língua diferente do velho português".

22 "com o fim de fixar a língua ou fazê-la voltar ao século XVI".

23 "A nossa gramática não pode ser inteiramente a mesma dos portugueses. As diferenciações regionais reclamam estilo e método diversos. A verdade é que, corrigindo-os, estamos de fato a mutilar ideias e sentimentos que nos são pessoais. Já não é a língua que apuramos, é o nosso espírito que sujeitamos a servilismo inexplicável. Falar diferentemente não é falar errado".

Beyond the idea of errors, the definition of dialect by Bluteau as a 'way of speaking a language in the provinces of the same kingdom, or conquests'[24] (Silva 1789, 435) was troubling. After all, it had been almost a century since Brazil was no longer part of the United Kingdom with Portugal, and the strong Lusophobe sentiment 'in a manifest or latent state, conveyed in biases and acts of hostility'[25] (Mendes 2010, 17) created a buzz in the attempts at linguistic approximations.

Dispelling judgments of the dialect's worth, but considering that one could still not speak of linguistic autonomy, Silvio Romero (1888) preferred the terms "Brazilian-Portuguese" or "Luso-Brazilian". He considered the term "dialect" to have three possible meanings: i. a synonym of language/idiom; ii. an inferior form of a language; and iii. a subdivision of an idiom. The first meaning was not applicable to the Portuguese of Brazil, since Brazilians and the Portuguese had a *common dialect*. He also dispelled the idea of inferiority, since in racial (and, by corollary, linguistic) hybridization,[26] he saw a potential nation (Schwarcz 1993). Hence, the third meaning remained: the subdivision of an idiom. However, "dialect" as a "subdivision" of an idiom was a future possibility for the Portuguese of Brazil: 'if we do not already have a completely marked dialect, we are going to have it',[27] since Brazilian is 'a new nationality, in addition to other elements that did not exist in the old metropolis, and a population is developing that can no longer be confused with the Portuguese people'[28] (Romero 1888, 311).

Other Brazilian intellectuals also defended that the Portuguese of Brazil could not receive attributes of incorrectness or ignorance, concepts that, at the time, were normally coupled with the term dialect. In fact, despite entitling his work *O dialecto brazileiro* 'The Brazilian dialect', Pacheco da Silva Júnior (1880) was against this designation, because he understood that some Portuguese intellectuals saw it as an inferior form of language (Romero 1888, 311).

A similar attitude was held by Nascentes (1922, ²1953), who emphasized the autonomy of the Portuguese of Brazil[29] and stressed that accepting the label of dialect depended on the concept given to it. After all, 'certain laws of Portuguese grammar have lost their application among us; we speak differently without meaning that we speak incorrectly, since we express ourselves according to the new paths that the Portu-

24 "modo de falar huma língua nas provincias do mesmo reino, ou conquistas".

25 "em estado manifesto ou de latência, veiculad[o] em preconceitos e actos de hostilidade".

26 Hybridization would be a form of achieving whiteness.

27 "si não temos já um *dialecto completamente acentuado*, marchamos para tel-o".

28 "uma nacionalidade nova, a que juntaram-se outros elementos que não existiam na velha metrópole, e vão formando um povo que se não póde mais confundir com o povo portuguez".

29 In 1924, Graça Aranha, upon addressing the Brazilian Dictionary, proposed the exclusion of words characteristic of the Portuguese world (Marques 1933, 75).

guese language has taken in Brazil'[30] (Nascentes 1922, [2]1953, 10–11). This position was assumed well before João Ribeiro (1933, [1]1887) and shared by Souza da Silveira ([9]1983, [1]1923), who also believed that using the term dialect depended on the concept: 'So, we speak a dialect? Not if we interpret dialect as synonymous with uneducated speech'[31] (Souza da Silveira ([9]1983, [1]1923). This, however, still assumed the reach of Lusitanian Portuguese through schools.[32]

We recall that models of racial analysis saw a problem of decadence in Brazilian miscegenation that, obviously, affected the linguistic issue. Therefore, the discourse on the depigmentation of skin was followed by that of the linguistic elimination of what was considered a reflection of the Brazilian social structure and, by extension, affected the "Brazilian dialect" designation, which held the prejudice of speech fraught with errors.

Revisiting the idea of Europe as a civilizing model brought about by the first world war (Chauí 2000) pushed back against racial theories, thus allowing for the emergence of a nationalist perception. In this context, the term dialect gave way to a reference of Portuguese of Brazil.

On the other side of the Atlantic, the idea of Portugal as a 'spatially deterritorialized Portuguese nation'[33] grew (Feldman-Bianco 1992). In this context, the nationalist current opened the doors of the Lisbon Geographical Society (1878) with the goal of studying the languages spoken in the Portuguese colonies. Works by Adolfo Coelho on creole languages emerged and, at the turn of the century, a study of Portuguese dialects by Leite de Vasconcelos. Both reserved the space of overseas dialect for spoken Portuguese in Brazil. Adolfo Coelho facilitated the opposition between the conservative variety ('use of the old metropolis')[34] used by educated people and the Brazilian dialect used by the general population.

Within the proposal by the Lisbon Geographical Society, which aimed to study the languages spoken in the colonies and established the "Colonial Portuguese Course" dedicated to teaching linguistic courses of Overseas Portuguese, Adolfo Coelho (1967, [1]1883) highlighted the "creolizing trend" of the spoken language in Brazil, namely through the suppression of plural forms, reduced to the determiner.

This interpretation was reinforced by Gonçalves Viana (1931), for whom syntactic constructions related to pronominal placement (proclisis in the main clause and enclisis in the dependent clause) and to Subject-Verb order in questions 'are not and have

30 "[c]ertas leis da gramática portuguesa perderam aplicação entre nós; falamos diferentemente sem que por isso falemos errado, pois nos exprimimos de acordo com os novos rumos por que marchou no Brasil a língua portuguesa".

31 "Falamos então um dialeto? Não, se interpretarmos dialeto como sinônimo de falar inculto".

32 This position continued into the following decades, with Chaves de Melo (1946, [4]1981), Monteiro (1931, [3]1959) and Elia (1979).

33 "nação portuguesa espacialmente desterritorializada".

34 "uso da antiga metrópole".

never been Portuguese; they are creoles, as most of the particular characteristics of Brazilian pronunciation are also creoles, diverging from those of Portugal'[35] (Viana 1931, 130).

Dedicated to studies of Portuguese dialects, Leite de Vasconcelos ([1]1901) listed a range of linguistic facts that, in his opinion, formed the profile of the "Brazilian dialect". Among the phonological aspects, he indicates front and back vowel lowering, the monophthongization of *ou* (*pôco*, *loco*), the dropping of the final r (*flô*, *ardê*), and iotization (*muyé* = *mulher*). On the morphological level, he points out the suppression of plural marking on nouns and the use of the diminutive on gerunds (*dormindinho*). On the syntactic level, he calls attention to enclisis in subordinate clauses, the use of nominative pronouns in the accusative function and the preposition *em* 'in' with movement verbs. Noting that the discussion on the dialect status of Portuguese of Brazil was on the agenda, he concluded that, just as the Portuguese of Trás-os-Montes had the quality of a dialect, so did that of Brazil.

Nonetheless, the "language bosses" (Lima Sobrinho 1977) in Portugal rejected the term "Brazilian dialect", since its recognition could lead to cultural erosion, the reason for which, in general, they insisted that Brazilians made mistakes which should be fully banned in order to pave the way towards a uniformization with Lusitanian uses. With respect to a common language, the "conflict" was intensified by the desire and, at the same time, the disregard for the classical linguistic model for Brazil. Júlio Moreira (1922) condemned the association between the educated Portuguese of Brazil with classical Portuguese and declared that 'Brazil should strive to not deviate from the language that is written in Portugal'[36] (Moreira 1922, 342). While the advent of the Republic in Brazil (1889) and in Portugal (1910) removed the last entanglement in the dynastic plot that insisted on Brazil joining the Lusitanian matrix, Portuguese grammarians continued to project the Lusitanian linguistic blanket over the tropics.

3.4 A break for Ruy

The end of the 19th century was dominated by the sanitation policy that resulted in the vaccine revolt in 1904. Transforming into linguistic purism, a new season of linguistic discussions opened up. On the occasion of the project for the first Brazilian Civil Code, Ruy Barbosa discussed grammatical issues with Carneiro Ribeiro (1950) (Pagotto 2011). In defense of philological prejudices, Ruy Barbosa condemned the Portuguese of Brazil for resulting from linguistic contact and from the assimilation of foreign loanwords. He characterized it as a "bastard language", a "promiscuous dialect" demand-

35 "não são nem foram nunca portuguesas; são crioulas, como crioulas são também as mais das particularidades de pronúncia brasileira, que das de Portugal se afastam".

36 "o Brasil deverá esforçar-se por não se desviar da língua que em Portugal se escreve".

ing a "sense of vernacularity", of "genuine native wording" (apud Freire 1921, X), understood as classical Portuguese, or as Lusitanian Portuguese, a position questioned by Mendonça (1936):

> 'On the one hand, we received the previously developed written language, which was brought here by writers and doctors. On the other, a language came to us spoken by the colonizers, which later with time became increasingly different from the original. But while this popular language evolved, freely developing its trends, the written language, the language of doctors remained intact, with no consideration for the language of the people. And all of this constituted a reversal of the order of things that Alencar and others aimed to correct, but that the reactionaries—especially the grand shaman (Ruy Barbosa)—fought to establish'[37] (Mendonça 1936, 305).

Linguistic tension permeates the extensive correspondence between Godofredo Rangel and Monteiro Lobato (Lobato 1944), consisting of distress by the latter due to the mythical figure of Ruy obeying the purist position and, consequently, the laws of Lusitanian grammar, which gained ground particularly in the working-class press (Cruz 2013).

Despite the widespread Luso-tropicalism in the 1940s, the conflict still remained between the two views, traditionalist and nationalist, appropriating 'old and new prejudices. On one side, exaggerated purism, a fossilized concept of language; on the other, the aspiration for a national language, separate from the Portuguese language,'[38] understood as the nation's expression of sovereignty (Cunha [2]1970, 15–16). From the 1930s onward, social evolutionism began to decline, and new nationalist whirlwinds brought into focus titles such as *Língua Nacional* ('National Language', Ribeiro 1933, [1]1887; Jucá Filho 1937), *Idioma Nacional* ('National Idiom', Nascentes 1928), *Cultura da Língua Nacional* ('Culture of the National Language', Marques 1933), *A Língua Vernácula* ('The Vernacular Language', Sá Nunes 1938), among others. This new way of referring to the Portuguese language, by leaving the language unnamed, 'has [...] the purpose of calming the nationalist zeal'[39] (Cunha [2]1970, 96) on both sides of the Atlantic, merely recognizing the existence of 'general trends characteristic of a national dialect'.[40]

---

37 "Recebemos por um lado a língua escrita, já trabalhada, que para cá foi trazida por cronistas e doutores. Por outro, chegou-nos uma língua falada pelos colonizadores, a qual depois com o tempo cada vez mais se diferenciou da primitiva. Mas enquanto esta língua popular evoluía, desenvolvendo livremente suas tendências, a língua escrita, a língua dos doutores permanecia intacta, sem tomar conhecimento sequer da língua do povo. E tudo isto constituía uma inversão da ordem das coisas que Alencar e outros procuraram corrigir, mas que os reacionários – sobretudo o grande pagé (Ruy Barbosa) – lutaram para restabelecer".

38 "velhos e novos preconceitos. De um lado o purismo exagerado, uma concepção fossilizada da língua; de outro, o anelo por uma língua nacional, própria, desvinculada da portuguesa".

39 "tem [...] o condão de acalmar os zelos nacionalistas".

40 "tendências gerais características de um dialeto nacional".

But there were those who went further and called it the *Língua Brasileira* ('Brazilian Language', Sanches 1940). After all, the issue of designating the spoken language in Brazil had become part of the legal world as well. In 1935, the City Council of Rio de Janeiro approved the Proposed Bill to accept the name *língua brasileira* ('Brazilian language', Cunha 1980); the National Traffic Code used the expression Brazilian language (Chaves de Melo 1946, [4]1981, 26) and article 35 of the Transitional Provisions of the Federal Constitution of 1946 determined the establishment of a 'committee to provide their opinion on the designation of the national idiom'[41] (Sousa da Silveira 1983, [1]1923).

Amid the discussions on assuming the Brazilian language label, Nascentes changed the focus and assumed a different perspective from the one he had expressed in the 1920s. Under the titles *Independência literária e unidade da língua* 'Literary independence and language unity' and *Língua Brasileira* 'Brazilian Language'[42], he began to defend the idea that the differences were not central, such that the spoken language in Brazil should be seen as a variant of the Lusitanian variety. Contrary to political and literary independence, which occurs through human actions, he argued that linguistic autonomy depends on natural processes. In an interview with the newspaper *O Globo* (29/7/1935), he defined the spoken language in Brazil as: 'A very distinctive dialect of the Portuguese language and nothing more than a dialect'[43] (Nascentes 2003, 314). The reversal by Nascentes is emblematic of the comings and goings of Brazilian intellectuals in relation to the infiltration of Portuguese ideas.

Almost in parallel to what Nascentes wrote, Paiva Boléo (1943) claimed that 'rigorously speaking, there is no Portuguese, French or English language [...] There are dialects, tongues and various languages within the same linguistic community'[44] (1943, 10). The notion of a broad linguistic community, including Portugal and Brazil, conforms with the discourse of building a deterritorialized Portuguese nation.

Despite claiming that the notion of language derives from the abstraction of these particularisms, the philologist indicated features that characterized spoken Portuguese in Brazil to show each was present in some village on the peninsula or on some Portuguese island and, therefore, belonged to the Portuguese domain: 'One may come to discover that many Brazilianisms were, after all, Lusitanian dialectalisms'[45] (Boléo 1943, 68). The following would be from Portuguese: the suppression of the "r" on

41 "comissão para opinar sobre a denominação do idioma nacional". Sousa da Silveira, as a correspondent, concluded that the national language of Brazil is the Portuguese Language ([9]1983, [1]1923, 291-293).

42 Despite the lack of dating on the texts, since they make note of facts from 1937, one can deduce that these texts are after this period.

43 "Um dialeto muito caracterizado da língua portuguesa e nada mais que um dialeto".

44 "rigorosamente falando, não existe uma língua portuguesa, francesa ou inglesa [...] O que há é dialectos, falares e linguagens diversas dentro da mesma comunidade linguística".

45 "É possível que se venha a descobrir que muitos brasileirismos eram afinal dialetalismos lusitanos".

verbs, the transition from "lh" to "i" in *muié* 'woman', *orvaio* 'dew' and the palatal affricate in *noite* 'night'.

Contrary to Leite de Vasconcelos, Gonçalves Viana and Renato Mendonça considered African influence to be the explanation for differences between Brazilian Portuguese and European Portuguese, while Paiva Boléo attributed them 'to a natural need to simplify, experienced by uneducated people in Brazil'[46] (1943, 32). Although he did not use term error, it is difficult to separate the 'natural need to simplify' of 'uneducated people' from the scope of error. He defended that the supposed Brazilian linguistic facts, with the exception of plural marking, 'do not affect the structure of the language, much less do they form a dialect'[47] (1943, 47). The Portuguese from Brazil could not be considered 'either a dialect or a language, but a 'tongue''[48], assuming, randomly, that 'a tongue is broader'[49] than a dialect (1943, 16).

The rhetorical diversion that established the language-tongue-dialect order in place of language-dialect-tongue strengthened and upgraded the Brazilian case, elevating it almost to the language category. However, being a tongue, Brazilian Portuguese was so close to European Portuguese that if one removed its typical linguistic features, one could conclude more of the same, which is that Brazilian Portuguese is 'just a variant of European Portuguese'[50] (1943, 10).

Despite the difficulty in establishing distinctions between dialect and tongue, the proposal by Paiva Boléo was accepted by Nascentes (1922, [2]1953), who claimed it was 'less inconvenient to call it a tongue than a dialect'[51] ([2]1953, 17) regarding Brazilian Portuguese.

On this side of the Atlantic, reviving the idea of language as an abstraction came with the *Introdução ao Estudo da Língua Portuguesa no Brasil*[52] 'Introduction to the Study of the Portuguese Language in Brazil', by Serafim da Silva Neto (1977, [1]1950), who maintained that, despite the current language of Brazil differing from that of Portugal, 'we have a common linguistic structure and [...] we are part of the same linguistic domain'[53] (1977, 20). It is within this linguistic domain that he identifies not only differences from the current language, but also from the common (or standard) language whose roots date back to the 16th and 17th centuries, thus revealing Silva Neto's sheer audacity (1977, 20).

---

46 "a uma necessidade natural de simplificação experimentada pelo povo inculto do Brasil".

47 "não afectam a estrutura da língua, e muito menos chegam para formar um dialecto".

48 "nem dialeto nem língua, mas 'falar'".

49 "o falar é que é mais amplo".

50 "apenas uma variante do português europeu".

51 "menos inconveniente em chamar falar do que em dialeto".

52 Despite the 1st edition having come out only in 1950, the author had been working on Brazilian Portuguese since 1934 (Silva Neto 1977, 15).

53 "constituímos uma estrutura linguística comum e [...] fazemos parte do mesmo domínio linguístico".

> 'We have to take into account the existence of an educated Brazilian dialect that lies within the Portuguese language, but is different from the overseas one. This means that the *common language*, which is a superior kind of expression, presents two varieties in the Portuguese linguistic domain: European and American.'[54]

The recognition of a standard Brazilian norm gains importance with the support of figures such as Cunha ([2]1970) and Houaiss ([3]1992). Considering the diplomacy of linguistic unity, the latter suggested that, by focusing on the differences, more weight was placed on them than should have been. At the same time, he insisted on the need to put an end to the linguistic policy that recognized a single grammar and imposed teaching the EP modality in Brazil. Cunha posited "idiomatic unity", without entailing a linguistic standardization in the scope of the norm, because 'if a language may embrace various systems [...], it also may admit various norms' (Cunha [2]1970, 80).[55] He thus distanced himself from the prescriptive norm that assumed linguistic standardization and, out of respect for linguistic plurality, proposed the descriptive norm, which, instrumentalized in teaching, would guarantee a 'superior unity of the Portuguese language.'[56]

Going against the grain, Elia (1979) insists on Lusitanian prudence for the Portuguese language in Brazil, considering that 'the grammar of the Portuguese language of Brazil is essentially the same as in Portugal'[57] (1979, 148).

3.5 The linguistic communion: Brazilian Portuguese

The multiple labels for the Portuguese of Brazil in its almost two hundred years of independence is the result of contradictions and ambiguities that permeated the numerous conflicts surrounding the legitimacy of designations that erased the connection to the European matrix. Despite the scientific approach to the phenomena taken by historical-comparative linguistics and dialectology, gaps and doubts remained in the interpretation of the Brazilian dialect designation, due to models of cultural analysis. Social evolutionism led to the dialect-error/ignorance equation and nationalism fueled the dialect-submission relation.

In the form of reciprocal adjustments, a neutral nomenclature was adopted. Expressions such as national language, national grammar, vernacular language and our idiom started to be used to refer to the spoken language in Brazil, as well as in Portu-

54 "Temos de levar em conta a existência de um dialeto culto brasileiro que, embora dentro da língua portuguesa, difere do de além-mar. Isso quer dizer que a *língua comum*, que é um tipo superior de expressão, apresenta, no domínio linguístico português duas variedades: a europeia e a americana".

55 "se uma língua pode abarcar vários sistemas [...] pode também admitir várias normas".

56 "unidade superior da língua portuguesa".

57 "a gramática da língua portuguesa do Brasil é essencialmente a mesma de Portugal".

gal. Then, the ideology of the Luso-Brazilian brotherhood in the 1930s regulated the designations with the help of a modifier (Portuguese of/in Brazil). A reference to Brazil was made, but the link to Portugal was maintained, since the latter variety was not identified with modifiers. This strategy still preserved a trace of the notion of subordination to the European matrix.

The final solution did not result from an ideological view, but from practical issues. As shown by Coelho/Silva (2018), syntactic studies within the Principles and Parameters Model invigorated comparisons between the languages or linguistic varieties. The opposition between the grammar of Portuguese from Brazil (PB) and the grammar of Portuguese from Portugal (PP) was created, as evidenced in Galves (1984). In the courses of Historical Linguistics at the State University of Campinas (Unicamp) in the 1980s, Mary Kato and Fernando Tarallo openly assumed that, due to the entanglement of labels and ideologies that were associated to them, they agreed to adopt Brazilian Portuguese (BP) in contrast to European Portuguese (EP). The research carried out in historical linguistics at Unicamp organized in the book *Português Brasileiro: uma viagem diacrônica* (Roberts/Kato 1993, 'Brazilian Portuguese: a diachronic journey') confirmed the name Brazilian Portuguese. It was no longer a subordination, but an institutionalization of differences. The triangle proposed by Eduardo Carlos Pereira (1916) was thus settled: the Portuguese language occupies the vertex (16th century) with European Portuguese and Brazilian Portuguese at the base angles.

4 A view of the question based on the notion of grammar

As we saw above, the debate on the question of the difference between Brazilian and European Portuguese was, to a great extent, ideological. The linguistic considerations have always been embedded in the issue of the norm, on the one hand, and, on the other, in the issue of the political and cultural separation of Brazil and Portugal. The overall assumption of the existence of a Brazilian language is more a political point of view than the consequence of accepting a change in the way the language that was brought to the colony evolved and acquired new features on all linguistic levels. Moreover, most of the innovations are considered by many to be a corruption of the Portuguese norm, which continues to be taken as the example to be followed. Finally, even if, on the level of style, the difference is acknowledged, the opinion of the majority is that 'the grammar of Brazilian Portuguese is essentially the same as in Portugal'[58] (Elia 1979, 148), because 'it has the same structure'. Given the tension established between the evidence of some differences and the possibility of proclaiming similarity,

58 "a gramática da língua portuguesa do Brasil é essencialmente a mesma de Portugal".

words fail to express the exact relationship between the two varieties. The issue, therefore, remains unresolved, in the absence of adequate concepts to help understand where the breaking point is to allow one to claim without a doubt that the two languages are just two different realizations of the same underlying abstract entity or that they have reached said breaking point.

We would like to suggest that a way to solve this apparent dead end, from the point of view of linguistic theory, lies in the dichotomy discussed by Chomsky (1986) between External Language (or E-Language) and Internal Language (or I-Language).[59] According to Chomsky (1986, 19), "Structural and descriptive linguistics, behavioral psychology and other contemporary approaches tended to view a language as a collection of actions, or utterances, or linguistic forms (words, sentences), paired with meanings, or as a system of linguistic forms or events". As an example, he takes Bloomfield's definition of language as "the totality of utterances that can be made in a speech community". This conception treats language as an external object, in the sense that it is external to speakers' minds. At the same time, it corresponds to an extensional definition, which consists of listing the elements that language in general and specific languages in particular are made of. By contrast, the notion of I-Language refers to the knowledge internalized by speakers, which enables them to produce and interpret an infinite number of utterances, and more.

> The system of knowledge attained—the I-language—assigns a status to every relevant physical event, say, every sound wave. Some are sentences with a definite meaning (literal, figurative, or whatever). Some are intelligible with, perhaps, a definite meaning, but are ill-formed in a way or another ("the child seems sleeping"; "to whom did you wonder what to give?" in some dialects; "who do you wonder to whom gave the book?" in all dialects). Some are well-formed but unintelligible. Some are assigned a phonetic representation but no more; they are identified as possible sentences of some language, but not mine. Some are mere noise. There are many possibilities. Different I-languages will assign status differently in each of these and other categories. The notion of E-language has no place in this picture (Chomsky 1986, 26).

In a nutshell, viewing languages as I-languages is a shift from looking at some physical output to considering the abstract mechanism that generates and interprets it. A specific language (or grammar) from this point of view corresponds to the parametrization of the language faculty, or Universal Grammar (UG). Two I-languages are different if they contain at least one parameter fixed in the opposite way. When this occurs, not only do the two grammars produce different utterances, but *they also assign different structures to superficially identical utterances*. From this point of view, the case of Brazilian and European Portuguese is particularly interesting because, due to their recent split, they still share a large part of their lexicon, and they have the same morphology and phonology. If we can show that, in spite of these similarities, the status assigned by

59 What follows is based on Galves (1998).

Brazilian and Portuguese speakers to identical utterances is different, we can justify the claim that the Brazilian grammar is distinct from the Portuguese grammar.

Shifting from E to I-language, questions about the difference change from whether the list of different words, different phonetic realizations, and different syntactic constructions in BP is sufficiently large to define a new language, to whether the knowledge internalized by Brazilian speakers (their competence) is the same as that of Portuguese speakers. Even if we know that the notion of Brazilian speakers and Portuguese speakers is in large part an abstraction, since a great deal of variation exists on both sides, we can rely on the facts that clearly characterize the two varieties as distinct in order to try to answer this question in an objective way.

To characterize the Brazilian I-Language, in contrast to the Portuguese I-language, i.e. what constitutes the distinctive competence of Brazilian speakers, we have three types of data at our disposal. The first one are the utterances produced in one community that, in the other, either are considered completely impossible or receive a different interpretation. This sort of argument is only superficially identical to the reasoning of the traditional approach based on E-language, since it is not only the existence versus absence of certain linguistic objects that is relevant but also the interpretation associated to these objects, i.e., the structure projected on them by the speakers. Moreover, it is not the linguistic object per se that is important but what it reveals of the grammar that generates it. The second type of data comes from frequency. Some constructions are possible in both EP and BP, but with different frequencies. This fact suggests that they do not have the same status in each grammar. Finally, and this is a contribution offered by the generative approach, the third type of data has to do with the different extension of restrictions operating on apparently identical constructions (for instance, the use of null subjects or null objects, as we will see below). The observation of this last group is more difficult and presupposes a theoretical model leading one to search for data that are not immediately given by intuitive comparison. Note that, in this case, the *intra-linguistic* interplay between possible/impossible sentences is as important for the argumentation as the *inter-linguistic* contrasts involving the existence versus absence of certain constructions. In the remainder of this section, we shall illustrate each class of phenomena and discuss their relevance to the issue at stake.

As for the first group of data, we refer the reader to Section 2 above, in which the main syntactic differences between EP and BP are presented. This first review, which excludes any normative consideration that may lead one to disregard some of the most important differences, suggests that important modifications occurred in Portuguese in Brazil (cf. also ↗6 Historical syntax), since some syntactic aspects differentiate it not only from EP but also from other Romance and Indo-European languages. In this section, we shall concentrate on questions of utterance interpretation.

Consider for instance the following sentence:

(14) O Pedro é difícil de pagar
the Pedro is difficult to pay
'It is difficult to pay Pedro'

In EP, the only possible interpretation for (14) is, like in English, that it is difficult to pay Pedro. In BP, (14) can have the same meaning, but another interpretation is available, in which *o Pedro* is no longer the object of *pagar* 'pay', but its subject, with the interpretation that it is difficult for Pedro to pay people. Note that nothing in the visible linguistic material of the sentence guarantees one interpretation or the other, since both the subject and the object of *pagar* lack morphological realization, and the NP *O Pedro* is, in any case, in the subject position of the main clause. The difference between EP and BP, in this case, is not in the utterance itself, but in the interpretation of the utterance, more specifically, the interpretation of the null arguments of the infinitive verb. This is an I-language difference. As such, we expect to find other expressions with this type of difference. This is the case for sentences like (15).

(15) O Dr. Silva operou esta manhã
the Dr. Silva operated this morning
a) 'Dr. Silva performed an operation this morning' (EPOK, BPOK)
b) 'Dr. Silva was operated on this morning' (EP*, BPOK)

In EP, the only interpretation is that Dr. Silva is the agent of the sentence and the patient is indefinite. In BP, again, another interpretation is available, in which Dr. Silva is the patient of the operation. In this case, (15) has a meaning analogous to a passive sentence. Like in the case of (14), what is at stake is the interpretation of the null object of the verb. Again, Brazilian and Portuguese speakers evidence different knowledge.

In light of these facts, it is plausible to think that the differences observed in the frequency with which similar phenomena in EP and BP occur are related to the fact that they correspond to different underlying structures. One case at play again concerns null objects and subjects. Consider the following sentences, which are well-formed both in BP and EP (ec means empty category):

(16) *ec* iremos à praia amanhã
ec go.FUT to-the beach tomorrow
'We shall go to the beach tomorrow'

(17) A Joana viu *ec* na praia ontem
the Joana saw *ec* on-the beach yesterday
'Joana saw her/him/it on the beach yesterday'

Sentences (16) and (17), respectively, illustrate null subject and null objects and distinguish both EP and BP from languages like French or English, in which no such null

categories can appear. However, many studies have shown that speakers of BP prefer to use lexical pronouns instead of null subjects (Galves 1987; Duarte 1995; among others and, from a diachronic point of view ↗6 Historical syntax). Conversely, for anaphoric objects, null objects are their preferred strategy (cf. Duarte 1986), and 3rd person object clitics are used mostly in written language. Obviously, different frequencies can have different explanations. As for 3rd person clitics, they are simply disappearing from natural use in Brazil (cf. ↗6 Historical syntax). In the case of null subjects, some researchers have suggested that BP is losing the null subject property as well (Duarte 1995). Other approaches to this question, however, provided evidence that the null subject property is still productive in BP, but that it has other characteristics. One of those characteristics is that it can receive a generic (or indeterminate) interpretation, as we saw above in Section 2, ex. (8), in which EP requires the presence of the clitic pronoun *se*. Comparative studies have shown that other null subject languages behave like BP, and differently from EP. Those languages have been called "Partial Null Subject languages". Therefore, we see that different frequencies can also be explained, to a great extent, by grammatical, or I-language differences. In fact, the lower rate of null subjects in BP derives from the fact that the structural contexts in which they can be interpreted as personal pronouns, i.e with a specific reference, are more restricted than in EP, since, contrary to the latter language, the former needs a local antecedent (cf. Figueiredo Silva 1996; and references in ↗6 Historical syntax).

Interestingly, a contrary situation occurs with respect to null objects, which requires a local antecedent in EP but not in BP. Raposo (1986) claimed that in sentences like (18), in which the null object is embedded in an adjunct clause (surrounded by brackets), its reference cannot be retrieved.

(18)	o	pirate	partiu	para	as	Caraibas
	the	pirate	left	to	the	Caribbean
	[depois	de	ter	guardado	*ec*	cuidadosamente]
	after	of	having	put away	*ec*	carefully

In BP, sentences like (18) are perfectly comprehensible, with the null object understood as referring to the discourse topic (in this case, for instance, treasure). This indicates once again that BP speakers and EP speakers do not have the same I-language. For the former, null objects and lexical pronouns behave alike, with no structural restrictions on their ability to retrieve their antecedent. For the latter, on the contrary, null objects, although possible in some contexts, are sensitive to the structure. They behave more like null categories involved in constructions like interrogatives, which cannot be separated from the interrogative phrase by adjunct clause boundaries, among others, as shown in (19):

(19) *O que o pirate partiu para as Caraibas
what the pirate left to the Caribbean
[depois de ter guardado *ec* cuidadosamente]?
after of having put away *ec* carefully

Parallel to what happens with null subjects, we can therefore say that whereas BP is a *Consistent Null Object* language, EP is a *Partial Null Object* language. Beyond terminology, this reinforces the conclusion drawn on the basis of the differences in interpretation observed for other constructions involving null arguments. If we define Language as I-language, we have strong evidence that the Brazilian I-language and the Portuguese I-language are different. This means that, while they share a large part of their lexicon, as well as their morphology and phonology, their internalized grammar is different. This is not a matter of degree as it would be from the point of view of E-language, but a matter of a distinct parameterization of Universal Grammar.

5 Concluding remarks

The debate and the controversy initiated in the 19th century are not over, which is witnessed by the very existence of this chapter. Still today, the question of the difference between the varieties of Portuguese on the two sides of the ocean is open, and is present at school, at university, and in the media. The term "variety" itself involves the idea of the fundamental unity defended by many scholars who claimed that there is only one language because "the structure is the same" (cf. Section 3). In Section 4, we argued that if we reason in terms of I-language, we show that BP and EP have different structures, in the sense that they correspond to different mental competences that attribute different structures to identical utterances. The fact remains that people have a strong feeling of identity due to the large overlap in lexicon and in the morphological part of the grammar, that is, words, which common sense believes is the core of language. Moreover, BP and EP have a common ancestor, which is sufficiently recent, that is fairly similar to both. On one side, José de Alencar reasoned as a linguist when he stressed the unavoidable split due to the geographical separation of the languages, and the strong effect of language contact in Brazil.

On the other side, this occurred without the power of written language and school, which, from the 19th century on,[60] strongly exercised a centripetal ideological force against the natural tendency for the languages to split. If languages are both bio-

60 On the issue of the diffusion of Portuguese in the colonial period, which is outside the scope of this chapter, we refer the reader to ↗2 The social history of Brazilian Portuguese.

logical and cultural entities, how can we possibly answer the question of their identity in a single way?

6 References

Alencar, José de (2006, [1]1865), *Iracema*, São Paulo, Ateliê Editorial.

Altman, Cristina (2004), *Filologia e Linguística—outra vez*, Filologia e Linguística Portuguesa 6, 161–198.

Andrade, Aroldo/Galves, Charlotte (2014), *A unified analysis for Subject Topics in Brazilian Portuguese*, Journal of Portuguese Linguistics 13/1, 117–147.

Antonelli, André (2011), *Sintaxe de posição do verbo e mudança gramatical na história do português europeu*, unpublished Doctoral thesis, Campinas, Unicamp.

Argote, Jeronymo Contador de (1725), *Regras da Lingua Portugueza, Espelho da Língua Latina*, Lisboa, Officina da Musica.

Avelar, Juanito/Galves, Charlotte (2011), *Tópico e concordância em português brasileiro e português europeu*, in: Amanda Costa/Isabel Falé/Pilar Barbosa (edd.), *Textos Selecionados: Actas do XXVI Encontro Nacional da Associação Portuguesa de Linguística (Porto, 2010)*, Lisboa, APL, 49–65.

Avelar, Juanito/Galves, Charlotte (2014), *O papel das línguas africanas na emergência do português brasileiro*, Linguística 30/2, 241–288.

Balbi, Adrien (1826), *Atlas ethnographique du Globe*, Paris, Chez Rey et Gravier.

Barbosa, Ruy (1949, [1]1902), *Obras completas de Ruy Barbosa*, vol. 39, tomo IV, Rio de Janeiro, Ministério da Educação e Saúde.

Barbosa, Ruy (1969, [1]1902), *Obras completas de Ruy Barbosa*, vol. 29, tomos I–III, Rio de Janeiro, Ministério da Educação e Saúde.

Binzer, Ina Von (1994), *Os meus romanos*, Rio de Janeiro, Paz e Terra.

Boléo, Manuel Paiva (1943), *Brasileirismos. Problemas de método*, Coimbra, Coimbra Editora Ltda.

Bonvini, Emilio (2008), *Línguas africanas e português falado no Brasil*, in: José Luis Fiorin/Margarida Petter (edd.), *África no Brasil. A formação da língua portuguesa*, São Paulo, Contexto, 15–62.

Carneiro Ribeiro, Ernesto (1950), *A Redação do Projeto do Código Civil e a réplica de dr. Ruy Barbosa: Tréplica*, Salvador, Aguilar e Souza.

Castilho, Ataliba (s/d), *Estudos linguísticos no Brasil. Notas para sua história*, https://periodicos.fclar.unesp.br/alfa/article/download/3198/2925, last accessed 16.02.2022.

Chauí, Marilena (2000), *Brasil. Mito fundador e sociedade autoritária*, São Paulo, Perseu Abramo.

Chaves de Melo, Gladstone (1946, [4]1981), *A Língua do Brasil*, Rio de Janeiro, Padrão (1ª ed. Rio de Janeiro, Agir).

Chomsky, Noam (1986), *Knowledge of language: its nature, origin, and use*, New York, Praeger.

Coelho, F. Adolfo (1967, [1]1883), *Os Dialectos Românicos ou Neo-Latinos na África, Ásia e América*, in: Jorge Morais-Barbosa (ed.), *Estudos Linguísticos Crioulos*, Reedição de artigos publicados no Boletim da Sociedade de Geografia de Lisboa, Introdução e Notas, Lisboa, Academia Internacional da Cultura Portuguesa, 1–234.

Coelho, Olga/Silva, Wellington (2018), *Páginas de história da terminologia relativa ao português brasileiro*, in: Ataliba Castilho (ed.), *O português brasileiro em seu contexto histórico*, São Paulo, Contexto, 72–95.

Costa, João (2004), *Subject positions and interfaces. The case of European Portuguese*, Berlin/New York, Mouton de Gruyter.

Costa, João (2010), *PB e PE: orientação para o discurso importa?*, Estudos da Língua(gem) 8/1, 123–143.

Cruz, Helena de Faria (2013), *São Paulo em papel e tinta: periodismo e vida urbana 1890–1915*, São Paulo, Arquivo do Estado de São Paulo.
Cunha, Celso ([2]1970), *Língua Portuguesa e realidade brasileira*, Rio de Janeiro, Tempo Brasileiro.
Cunha, Celso (1980), *Presença de Antenor Nascentes*, Studia 10, 149–169.
Dias, Gonçalves (1921), *Carta ao dr Pedro Nunes Leal*, Estante Clássica da Revista da língua Portuguesa 7, 127–132, reprint in: Pinto, Edith Pimentel (ed.) (1978), *O Português do Brasil: textos críticos e teóricos 2; 1820–1920. Fontes para a teoria e história*, Rio de Janeiro, Livros técnicos e científicos/São Paulo, Edusp, 33–38.
Duarte, Maria Eugênia (1986), *Variação e sintaxe: clítico acusativo, pronome lexical e categoria vazia no português do Brasil*, unpublished Master's thesis, São Paulo, PUC-São Paulo.
Duarte, Maria Eugênia (1995), *A perda do princípio "evite pronome" no português brasileiro*, unpublished Doctoral thesis, Campinas, Unicamp.
Dubert, Francisco/Galves, Charlotte (2016), *Galician and Portuguese*, in: Adam Ledgeway/Martin Maiden (edd.), *The Oxford Guide to Romance Languages*, Oxford, Oxford University Press, 411–446.
Elia, Sílvio (1979), *A unidade linguística do Brasil*, Rio de Janeiro, Padrão.
Elias, Norbert (1990), *O processo civilizador*, Rio de Janeiro, Jorge Zahar Editor.
Faraco, Carlos Alberto (2018), *Breve retrospectiva do pensamento linguístico-histórico no Brasil*, in: Ataliba Castilho (ed.), *O português brasileiro em seu contexto histórico*, São Paulo, Contexto, 32–71.
Feldman-Bianco, Bela (1992), *Saudade, imigração e a construção de uma nação (portuguesa) desterritorializada*, Revista Brasileira de Estudos Populacionais 9/1, 35–49.
Figueiredo Silva, Maria Cristina (1996), *A posição sujeito em português brasileiro—frases finitas e infinitivas,* Campinas, Editora da Unicamp.
Fiorin, José Luiz (2016), *Identidade nacional e exclusão racial*, Cadernos de Estudos Linguísticos 58/1, 63–75.
Freire, Laudelino (1921), *Gallicismos*, Rio de Janeiro, Tipografia Fluminense.
Freyre, Gilberto (1933), *Casa-grande & senzala: formação da família brasileira sob o regimen de economia patriarchal*, Rio de Janeiro, Maia & Schmidt.
Frota, Sónia/Vigário, Marina (2001), *On the correlates of rhythmic distinctions: the European/Brazilian Portuguese case*, Probus 13, 247–275.
Galves, Charlotte (1984), *Pronomes e categorias vazias em Português do Brasil*, Cadernos de Estudos Linguísticos 7, 107–136.
Galves, Charlotte (1987), *A sintaxe do português brasileiro*, Estudos Linguísticos 13, 31–50.
Galves, Charlotte (1998), *A Gramática do Português Brasileiro*, Língua e Instrumentos Linguísticos 1, 79–96.
Galves, Charlotte (2012), *Concordância e origens do português brasileiro*, in: Adeilson Sedrins et al. (edd.), *Por amor à Linguística: miscelânea de estudos linguísticos dedicados à Maria Denilda Moura*, Maceió, EDUFAL, 123–149.
Galves, Charlotte/Andrade, Aroldo/Faria, Pablo (2017), *The Tycho Brahe Corpus of Historical Portuguese*, http://www.tycho.iel.unicamp.br/~tycho/corpus/, last accessed 16.02.2022.
Galves, Charlotte/Britto, Helena/Paixão de Sousa, Maria Clara (2005), *The change in clitic placement from Classical to Modern European Portuguese: results from the Tycho Brahe Corpus*, Journal of Portuguese Linguistics 4/1, 39–68.
Galves, Charlotte/Gibrail, Alba (2018), *Subject inversion in transitive sentences from Classical to Modern European Portuguese: a corpus-based study*, in: Adriana Cardoso/Ana M. Martins (edd.), *Word Order Change*, Oxford, Oxford University Press, 163–178.

Galves, Charlotte/Kroch, Anthony (2016), *Main syntactic changes from a Principles-and-Parameters view*, in: Leo Wetzels/Sérgio Menuzzi/João Costa (edd.), *The Handbook of Portuguese Linguistics*, New York, Wiley & Sons, 487–503.

Galves, Charlotte/Paixão de Sousa, M. C. (2017), *The position of the verb in the history of Portuguese. Subject position, clitic placement and prosody*, Language 93/3, 152–180.

Galves, Charlotte/Torres Moraes, M. Aparecida/Ribeiro, Ilza (2005), *Syntax and morphology in the placement of clitics in European and Brazilian Portuguese*, Journal of Portuguese Linguistics 4/2, 143–177.

Geertz, Clifford (1989), *A interpretação das culturas*, Rio de Janeiro, Guanabara.

Guy, Gregory (1981), *Linguistic variation in Brazilian Portuguese: aspects of phonology, syntax and language history*, unpublished Doctoral thesis, Philadelphia (PA), University of Pennsylvania.

Hobsbawn, Eric (1990, [2]2008), *Nações e nacionalismo de 1780: programa, mito e realidade*, Rio de Janeiro, Paz e Terra.

Holm, John (2004), *Languages in contact: the partial restructuring of vernaculars*, Cambridge, Cambridge University Press.

Holmberg, Anders/Nayudu, Aarti/Sheehan, Michelle (2009), *Three partial null subject languages: a comparison of Brazilian Portuguese, Finnish and Marathi*, Studia Linguistica 63/1, 59–97.

Houaiss, Antônio ([3]1992), *O Português no Brasil*, Rio de Janeiro, Revan.

Jucá Filho, Cândido (1937), *Língua Nacional: as diferenciações entre o português de Portugal e o do Brasil autorizam a existência de um ramo dialetal do português peninsular?*, Rio de Janeiro, Estabelecimento Gráphico Apollo.

Kabatek, Johannes (2015), *Sobre usos y abusos de la terminología lingüística*, Revue de Linguistique Romane 79/315–316, 331–359.

Kabatek, Johannes/Wall, Albert (edd.) (2013), *New Perspectives on Bare Noun Phrases in Romance and Beyond*, Amsterdam/Philadelphia, Benjamins.

Kato, Mary/Peres, João (edd.) (2005), *Studies in the comparative syntax of European and Brazilian Portuguese*, Journal of Portuguese Linguistics 4/2.

Lessa, Carlos (2008), *Nação e nacionalismo a partir da experiência brasileira*, Estudos Avançados 22/62, 237–256.

Lima Sobrinho, Alexandre José Barbosa (1977, [1]1958), *A Língua Portuguesa e a Unidade do Brasil*, Rio de Janeiro, José Olympio.

Lobato, Monteiro (1944), A barca de Gleyre, São Paulo, Companhia Editora Nacional.

Lombroso, Cesare (1876, [3]1884), *L'uomo delinquente: in rapporto all'antropologia, giurisprudenza ed alle discipline carcerarie*, Torino, Fratelli Bocca, https://www.lexml.gov.br/urn/urn:lex:br:rede.virtual.bibliotecas:livro:1884;000096760, last accessed 16.02.2022.

Lucchesi, Dante (2012), *A deriva secular na formação do português brasileiro: uma visão crítica*, in: Tânia Lobo et al. (edd.), *ROSAE: linguística histórica, história das línguas e outras histórias*, Salvador, EDUFBA, 249–274.

Lucchesi, Dante/Baxter, Alan/Ribeiro, Ilza (edd.) (2009), *O Português Afro- Brasileiro*, Salvador, Edufba.

Magalhães, Domingos Gonçalves de ([3]1864), *A Confederação dos Tamoyos*, Rio de Janeiro, Garnier, https://digital.bbm.usp.br/bitstream/bbm/4154/1/010874_COMPLETO.pdf, last accessed 16.02.2022.

Marques, Xavier (1933), *Cultura da Língua Nacional*, Bahia, Escola de aprendizes artífices.

Martins, Ana Maria (1994), *A história dos clíticos em português*, unpublished Doctoral thesis, Lisboa, Universidade de Lisboa.

Mendes, José Sacchetta (2010), *Laços de sangue, Privilégios e Intolerância à Imigração Portuguesa no Brasil (1822–1945)*, Porto, Cepese.

Mendonça, Renato (1936), *O Português do Brasil*, Rio de Janeiro, Civilização Brasileira S.A.

Monteiro, Clóvis (1931, [3]1959) *Português da Europa e Português da América*. Rio de Janeiro, Livraria Acadêmica.
Moreira, Julio (1922), *Estudos da Lingua Portuguesa. Subsidios para a Syntaxe Historica e Popular*, Lisboa, Livraria Classica Editora de A.M. Teixeira.
Munhoz, Ana/Naves, Rozana (2012) *Construções de Tópico-Sujeito: uma proposta em termos de estrutura argumental e de transferência de traços de C*, Signum. Estudos de Linguagem 15, 245–265.
Naro, Anthony/Scherre, Marta (2007), *As origens do português brasileiro*, São
Paulo, Parábola Editorial.
Nascentes, Antenor (1922, [2]1953), *O linguajar carioca*, Rio de Janeiro, Edição da Organização Simões.
Nascentes, Antenor (1928), *O idioma nacional*, São Paulo, Companhia Editora Nacional.
Nascentes, Antenor (2003), *Estudos dialetológicos*, Rio de Janeiro, Academia Brasileira de Letras.
Negrão, Esmeralda/Viotti, Evani (2008), *Estratégias de impessoalização no português brasileiro*, in: José Luiz Fiorin/Margarida Petter (edd.), *África no Brasil, a formação da língua portuguesa*, São Paulo, Contexto, 179–203.
Noll, Volker (2008), *O português brasileiro, formação e contrastes*, São Paulo, Globo.
Nunes, Jairo (1990), *O Famigerado SE: uma análise sincrônica e diacrônica das construções com SE apassivador e indeterminador*, unpublished Master Thesis, Unicamp.
Oliveira, Marilza de (2019), *A colocação pronominal como pista de "grupos estabelecidos e outsiders"*, in: Dermeval da Hora et al. (edd.) *Estudos Linguísticos, Teorias e Aplicações. Contribuição da ALFAL*, São Paulo, Terractoa, 97–114.
Oliveira, Marilza de (2021), *Ênclise pronominal na fatura linguística do Império e da República*, in: Jânia Ramos/Marilza de Oliveira (edd.), *Dialetação e povoamento: da história linguística à história social,* São Paulo, Contexto.
Pacheco da Silva Jr., Manuel (1880), *O dialecto brasileiro*, Revista Brasileira 5, 487–495.
Pagotto, Emílio (2007), *Crioulo sim, crioulo não, uma agenda de problemas*, in: Ataliba Castilho et al. (edd.), *Português Brasileiro: descrição, história e aquisição*, Campinas, Pontes, 461–482.
Pagotto, Emílio (2011), *Ruy Barbosa e a crise normativa brasileira*, in: Dinah Callou/Afrânio Barbosa (edd.), *A norma brasileira em construção: cartas a Rui Barbosa (1866 a 1899)*, Rio de Janeiro, Fundação Casa de Rui Barbosa, https://pt.scribd.com/document/362930390/Rui-Barbosa-e-a-Crise-Normativa-Brasileira, last accessed 16.02.2022.
Paixão de Sousa, Maria Clara (2004), *Língua barroca: Sintaxe e história do português nos 1600*, unpublished Doctoral thesis, Campinas, Unicamp.
Pereira, Eduardo Carlos (1916), *Grammatica Historica*, São Paulo, Companhia Editora Nacional.
Peres, João (2013), *Semântica do sintagma nominal*, in: Eduardo Raposo et al. (edd.), *Gramática do Português*, vol. 1, Lisboa, Fundação Calouste Gulbenkian, 735–815.
Pontes, Eunice (1987), *O tópico no português do Brasil*, Campinas, Editora Pontes.
Proença, M. Cavalcanti (1965), *Iracema—Lenda do Ceará*, Rio de Janeiro, José Olympio.
Raposo, Eduardo (1986), *On the null object construction in European Portuguese*, in: Oswald Jaeggli/Carmen Silva-Corvalán (edd.), *Studies in Romance Linguistics*, Dordrecht, Foris, 373–390.
Révah, Ismael (1956), *L'évolution de la prononciation au Portugal et au Brésil du 16ème siècle à nos jours*, in: *Anais do primeiro congresso brasileiro da língua falada no teatro*, Salvador, Ministério da Educação e Cultura, 387–402.
Ribeiro, João (1933, [1]1887), *Grammatica portugueza*, Rio de Janeiro, Livraria Francisco Alves.
Ribeiro, João (1979), *A língua nacional e outros estudos linguísticos*, Petrópolis/Aracaju, Vozes/Governo do Estado de Sergipe.
Ribeiro, Priscila (2015), *Língua e sociedade: a ordem na escola republica paulista*, unpublished Doctoral thesis, São Paulo, University of São Paulo.
Roberts, Ian/Kato, Mary (1993), *Português Brasileiro: uma viagem diacrônica*, Campinas, Pontes.

Romero, Silvio (1888), *Estudos sobre a poesia popular do Brasil*, Rio de Janeiro, Laemmert, https://digital.bbm.usp.br/view/?45000008916&bbm/6851#page/314/mode/2up, last accessed 16.02.2022.

Sanches, Edgard (1940), *A língua brasileira*, São Paulo, Companhia Editora Nacional.

Sá Nunes, José (1938), *A língua vernácula*, Ed. Globo.

Sandalo, Filomena, et al. (2006), *Secondary stress in two varieties of Portuguese and the Sotaq optimality based computer program*, Probus 18/1, 97–125.

Scherre, Marta (1988), *Reanálise da concordância de número em português*, unpublished Doctoral thesis, Rio de Janeiro, Federal University of Rio de Janeiro.

Schmitt, Cristina/Galves, Charlotte (2016), *Bare nominals in Brazilian and European Portuguese: a historical perspective*, talk at Nominals, GETEGRA International workshop, Recife, 17–18.

Schwarcz, Lilia (1993), *O espetáculo das raças*, São Paulo, Companhia das Letras.

Schwarcz, Lilia (1998), *As barbas do imperador*, São Paulo, Companhia das Letras.

Silva, António de Morais (1789), *Diccionario da Lingua Portugueza*, composto por Rafael Bluteau. Reformado e acrescentado, Lisboa, Ferreira.

Silva Neto, Serafim da ([4]1977, [1]1950), *Introdução ao estudo da língua portuguesa no Brasil*, Rio de Janeiro, INL-MEC.

Souza da Silveira, Álvaro ([9]1983, [1]1923), *Lições de português*, Rio de Janeiro, Presença.

Souza, Camila Duarte (2014), *Eu te amo, eu lhe adoro, eu quero você: a variação das formas de acusativo de 2ª pessoa em cartas pessoais (1880–1980)*, unpublished Master' thesis, Rio de Janeiro, Federal University of Rio de Janeiro.

Tarallo, Fernando (1983), *Relativization Strategies in Brazilian Portuguese*, unpublished Doctoral thesis, Philadelphia (PA), University of Pennsylvania.

Tarallo, Fernando (1993a), *Diagnosticando uma gramática brasileira: o português d'aquém e d'além-mar ao final do século XIX*, in: Ian Roberts/Mary Kato (edd.), *Português Brasileiro. Uma viagem diacrônica*, Campinas, Editora Unicamp, 69–105.

Tarallo, Fernando (1993b), *Sobre a alegada origem crioula do português brasileiro: Mudanças sintáticas aleatórias*, in: Ian Roberts/Mary Kato (edd.), *Português brasileiro, uma viagem diacrônica*, Campinas, Editora da Unicamp, 35–68.

Teyssier, Paul (1976), *Manuel de langue portugaise (Portugal–Brésil)*, Paris, Klincksieck.

Teyssier, Paul (1980), *Histoire de la langue portugaise*, Paris, PUF.

Vasconcelos, J. Leite de (1970, [1]1901), *Esquisse d'une dialectologie portugaise*, Lisboa, Centro de Estudos Filológicos.

Vázquez Cuesta, Pilar/Mendes da Luz, Maria Albertina ([3]1971), *Gramática portuguesa*, 2 vol., Madrid, Gredos.

Viana, A. R. Gonçalves (1931), *Palestras filológicas*, Lisboa, Livraria Clássica Editora.

Vieira, Sílvia/Brandão, Sílvia (2014), *Tipologia de regras linguísticas e estatuto das variedades/línguas: a concordância em português*, Linguística 30, 81–112.

Vanderci de Andrade Aguilera

9 Dialectology and linguistic geography

Abstract: This article presents a historical panorama of dialectology and geolinguistics in Brazil and contains five sections: (1) an introduction offering an overview of the methods used in dialectology, such as vocabularies, monographs, and atlases; (2) the phases of the history of dialectology in Brazil, which initiated in 1826 and developed for almost two centuries culminating in the publication of the *Atlas Linguístico do Brasil* in 2014; (3) the regional history of geolinguistics in Brazil and its specific centers of study; (4) a discussion of some phonetic and lexical aspects registered in the *Atlas linguístico do Brasil*, comparing them to the proposal of the classical dialectical division by Nascentes (1953) and, finally, (5) some concluding remarks.

Keywords: Brazilian Portuguese, dialectology, geolinguistics, linguistic atlases, history of Brazilian Portuguese dialectology

1 Introduction

Dialectology, conceived as a systematic study of all forms of dialects, especially regional dialects, makes use of the three methods of dialectal studies: the compilation and analysis of (i) vocabularies, glossaries or dictionaries, (ii) monographs, and (iii) linguistic atlases. The history of dialectology in Brazil has not been different, as the studies of diatopic variation have been developing and gaining a considerable impulse, since the second half of the 20th century, and especially in the past 15 years.

Considering the diverse collection of Brazilian dialectal studies, some based on traditional dialectology and some associated with other theories, and the principles of variationist sociolinguistics, we decided to consider works of the following nature for our account: (i) works of lexicographic nature of regional variants before the implementation of dialectology in Brazil, (ii) pre- and post-dialectology monographs focusing on diatopic variations, (iii) studies based on traditional geolinguistics or pluridimensional geolinguistics which present linguistic maps as a result.

Taking the intervention of the Visconde de Pedra Branca from the work of Adrien Balbi (1826) as a starting point and continuing until the publication of the first two volumes of the *Atlas Linguístico do Brasil* (Cardoso et al. 2014), we have before us almost two centuries of dialectological studies.

https://doi.org/10.1515/9783110405958-010

2 Phases in the history of dialectology in Brazil

The literature indicates that already in the beginning of the 20th century, before a relatively productive panorama of studies about the Brazilian language and its dialects, the division of the history of dialectology in Brazil was one of the concerns of Nascentes, a renowned philologist and dialectologist from Rio de Janeiro, as demonstrated in his article *Études dialectologiques au Brésil*:

> "On peut diviser l'histoire des études dialectologiques au Brésil en deux phases : la première, de 1826, année dans laquelle le brésilien Borges de Barros fait paraître une étude dans le livre d'Adrien Balbi, jusqu'à 1920, année de la publication du livre, *O dialeto caipira*, de Amadeu Amaral ; la deuxième, de 1920 à nos jours. La première se caractérise par la prédominance des glossaires sur les études grammaticales; la deuxième, au contraire, para la prédominance des études grammaticales sur les glossaires" (Nascentes 1952, 181).[1]

Nascentes highlights the participation of Domingos Borges de Barros and Visconde de Pedra Branca in chapter five of *Introduction à l'Atlas ethnographique du globe*, written by the geographer Balbi. It presents general comments by Pedra Branca about the Portuguese transplanted to Brazil, emphasizing "the sweetness of the weather and the character of its inhabitants" which made it possible for this language to attain gentleness and pleasantness while preserving its original energy. Nascentes (1952, 181) adds that in the second part of the article, the Visconde compiles the Portuguese words[2] which changed their meaning completely in Brazil, like *faceira* (a woman with a big jaw in Portugal; a vain woman in Brazil) and some other nonexistent words in European Portuguese which were taken from the native or African languages by BP (Brazilian Portuguese), such as *quitanda* 'sweets' or 'the place where they are sold', *balaio* 'hamper', *cochilar* 'take a nap/to doze', *caçula* 'the youngest child', and *cipoada* 'a stroke, using a liana'.

This first phase saw the publication of national and regional lexicographic studies, often influenced by nationalist and jingoistic movements, and lasted almost one century. Nascentes (1952) provides a thorough and chronological account of this period, detailing the diverse formats and media used: newspapers articles, scientific magazines and book chapters, always including in the title of the work the adjective *brasileiro*. Along with general works, the author lists regional vocabularies and glossaries: *sul-rio-grandense*, *paraense, mineiro, pernambucano, cearense, baiano,*

1 'One could divide the history of dialectal studies in Brazil into two phases: the first one, beginning in 1826, when the Brazilian Borges de Barros published a study in Adrien Balbi's book, until 1920, the year of the publication of Amadeu Amaral's book *O dialeto caipira*. The second one from 1920 until today. The first one is characterized by a predominance in the production of glossaries over grammatical studies, while in the second one, in contrast, grammatical studies predominate over glossaries.'

2 The examples do not appear in Nascentes' article (1952), but they were collected by the authors from Balbi's work (1826, 172–175).

acreano, alagoano, paranaense, additionally emphasizing that African and Tupi contributions were the topic of several works during this phase.

Finally, the author exposes that Amaral had already in 1916 presented in Revista do Brasil the premises from *O dialeto caipira*, a work "qui a changé la direction des études dialectologiques au Brésil et leur a fait prendre une allure scientifique" (Nascentes 1952, 184).[3] Amaral's work (1953; [1]1920), as we have seen, closes one phase and starts another which extends until 1963, the year of the publication of the *Atlas prévio dos falares baianos* (APFB) by Rossi and his team.

Castilho (1972–1973) provides an account of the history, theory and the methods used in dialectological research. He recognizes that the dialectological studies in Brazil have their turning points between 1953 (opening of the *Centro de Estudos de Dialetologia Brasileira* in Rio de Janeiro) and 1958 (1st Brazilian Congress of Dialectology and Ethnography, in Porto Alegre). The author praises Serafim da Silva Neto as a great propagator of dialectology ideals and someone who did everything in his power to implement a "dialectological mentality" in the country, including the foundation of the *Revista Brasileira de Filologia*. He concludes that the best answer to the works of Silva Neto was given by Nelson Rossi, who in 1963 accomplished the *Atlas de Sergipe*, still unpublished (Castilho 1972, 120).

Two decades later, Ferreira and Cardoso (1994, 37–59) continue the history of dialectology in Brazil, proposing a three-phase division: the first from 1826 to 1920, the second from 1920 to 1952 and the third from 1952 until 1994, the year of their publication. About the first phase, as Nascentes and Castilho did, the authors mention the participation of the Visconde de Pedra Branca in Balbi's book (1826), and they list the same lexicographic works[4] that can be found in Nascentes' article (1952). As to the second phase, initiated by O *dialeto caipira*, Ferreira and Cardoso clarify that:

> 'The empirical knowledge of the linguistic reality and the lack of systematic fieldwork, which characterize the production of the first phase, remain as a trace of the second phase, even though a direct observation of the described area and the concern with a methodology that focuses on an examination of the considered reality in its different aspects is notable.' (ADD PAGE)[5]

The most significant works of this second phase are of monographic nature—systematic descriptions of the grammatical aspects of a dialect or regional variety: descriptions of the rural *caipira* dialect by Amaral (1920; 1976; [1]1920), the *carioca* variety by Nascentes (1953), *Nordestino* by Marroquim (1996; [1]1934), and *mineiro* (1938) and *goia-*

3 'which changed the direction of dialectological studies in Brazil and gave a scientific aspect to it.'

4 Ferreira and Cardoso (1994, 38–39) mention only 14 of the 24 works listed and commented on by Nascentes (1952).

5 "O conhecimento empírico da realidade lingüística e a ausência de trabalho de campo sistemático, que marcam a produção da primeira fase, permanecem como traço da segunda fase, ainda que se experimente a observação direta à área a descrever-se e a preocupação com uma metodologia de abordagem voltada para o exame da realidade considerada nos seus diferentes aspectos."

no (1944) by Teixeira. Besides these five works focusing on the description of a specific dialect (or variety), Ferreira and Cardoso (1994) emphasize other contributions: regional lexicons and glossaries, works of general character analyzing Brazilian Portuguese in a broad and globalized perspective, as well as studies on the African contribution to BP.

Ferreira and Cardoso point out that *O dialeto caipira* has the merit of being the pioneer in the description of the many aspects of a dialect, in the case of the *paulista* countryside *Além da Introdução* dialect, which presents, among other things, recommendations for the collection of dialectal data. The work comprises the following sections: Phonetics, Lexicology, Morphology, Syntax, and Vocabulary. The latter is given more emphasis in the work, for out of 192 pages, 111 are about the *caipira* vocabulary.

In this work, Amaral (1976, 42) recognizes that at the end of the 19th century, there was a well pronounced dialect, in the territory of the old province of S. Paulo, which gave the *Paulistas* the reputation of corrupting the vernacular with many and ugly language vices. The deep social and cultural changes that were occurring in that territory affected the *caipira* dialect in such way that it became increasingly rare and was doomed to disappear in a relatively short period of time. Nearly one century later, the *caipira* dialect is still very healthy in several parts of the Southeast, as well as in the Central-West and South of Brazil. Fortunately, Amaral was mistaken.

In the 1953 edition of *O linguajar carioca*, Nascentes presents information about the formation of the country and the evolution of the Portuguese language in Brazil for the first time. In the second chapter, there is a proposal for a dialectal division in Brazil and a discussion about other similar divisions. Nascentes knew the Brazilian territory from North to South, which gave him assurance of a proposal confirmed by the reality of facts. *O linguajar carioca* presents a detailed description of phonetic, morphological and syntactic features, as well as nearly 800 vocabulary entries.

Map 1, adapted from Nascentes (1953, 17) presents the proposal of his dialectal division according to the realization of pre-tonic vowels. The division basically distinguishes between a northern and a southern group of dialects and includes a dialectally "uncharacterized territory". The northern dialects are *amazônico* and *nordestino*, the remaining ones are considered to be part of the southern group.

Map 1: Map of Nascentes (1953, adapted by Kika Milani).

Marroquim (1996), in turn, addresses in *A língua do Nordeste* the popular languages of two states together: Alagoas and Pernambuco, which represent the two populations under only one dialectal aspect. He recognizes that the historical and ethnical formation of the *alagoanos* and *pernambucanos* is the same and their linguistic orientation is identical.

O falar mineiro, an article by Teixeira (1938), follows the same structure as the previous works. In the introduction, Teixeira justifies his own work:

> 'Amadeu Amaral was a pioneer in dialectological studies, in 1920, with his Dialeto Caipira. It was necessary a *paulista* to overcome the aggressive carranca (sic) of the old philology. Antenor Nascentes followed his footsteps in Linguajar Carioca. Mario Marroquim sang the language of Pernambuco and Alagoas with great proficiency. Minas kept quiet. When the winds of the philological nationalism sweeps through Brazil. Avariciously keeping other language gold mines. As a Brazilian, I feel a revitalizing impulse. As a *mineiro*, I like singing my homeland language, which I learnt in my wistful childhood' (Teixeira 1938, 4).[6]

The fifth and final monographic work that Ferreira and Cardoso (1994) include in the second phase is by Teixeira (1944) about the language from Goiás. Teixeira introduces a chapter about dialectological concepts, dividing it in three phases: the first, until 1835, in which the Portuguese tradition of formulating vocabularies and glossaries is kept; the second, already influenced by the ideals of Romanticism with a nationalist character. The beginning of the third phase, according to him, we undeniably own to the Modernist Movement, which valued what was from home, and awakened the interest of new generations to it. The third phase, according to the author, begins with *O dialeto caipira*. The description of the variety from Goiás follows the model of previous works and ends with a regional glossary with 250 entries, the format of which is different from those of Amaral (1920) and Nascentes (1953). This glossary has the following structure: classification (neologism, brazilianism, Portuguese); entry, in its popular form; grammatical category, meaning, dictionarization, accreditation, or the author's comments, and a collection of the variants.

In Table 1 we present two examples taken from Teixeira (1944):

Table 1: Examples from Teixeira's Glossary (1944).

classification	word		meaning		site
Pop. etym.	*Aprofia*	sf	debate, quarrel	altered form of *porfia*	Bela Vista
Brazilian.	*Saraguá*	sf	sickle w/o turn		Itaberaí

Teixeira's greatest merit, besides having described the *mineiro* and *goiano* dialects, was of being the precursor to geolinguistics in Brazil. Ignored by dialectologists until recently, Teixeira's work nonetheless preceded the proposal of a linguistic atlas of Brazil, established in the Decree of 1952, by eight years. Including five maps about

6 "Amadeu Amaral foi bandeirante dos estudos dialetologicos. Em 1920, com seu Dialeto Caipira. Era necessário um paulista pra vencer a carrança (sic) agressiva da velha filologia. Antenor Nascentes seguiu suas pegadas com Linguajar Carioca. Com grande proficiencia Mario Marroquim cantou a lingua de Pernambuco e Alagôas. Minas ficou quieta. Quando o sopro do nacionalismo filológico varre o Brasil afóra. Guardando avaramente outras minas de ouro de linguagem. Brasileiro sinto impulso de renovação. Mineiro gosto de cantar a lingua de minha terra. Que aprendi na infancia saudosa."

phonetic phenomena recurrent in the variety of Goiás, it preceded the APFB by almost 20 years. On page 61, for example, Teixeira (1944) presents the *Carta do L final*, which consists of a draft of a map (cartogram) from southern Goiás, including on its upper left side map keys with variants, displaying apocope (*siná* 'signal'), intensification (*sinali*), conservation (*sinal*) and change for r (*sur* 'south'). These variants are placed on 18 points on the map. When there are two or more variants per point, the rectangle is subdivided, showing the various forms collected.

We believe that the periodization of the history of dialectology could be revised and that Teixeira's work about the *goiano* dialect should mark the beginning of the third phase, in which the theoretical discussions and methodologies, which will guide the next generations of geolinguists in Brazil, were published. The first drafts of maps introduced by Teixeira (1944) are uncontested proof of this effort.

According to Ferreira and Cardoso (1994) and other scholars, however, the landmark of the third phase is the Decree-law no. 30 643, from March 20, 1952. In the third article of the decree, the task of coordinating the linguistic atlas of Brazil is assigned to the *Comissão de Filologia da Casa de Rui Barbosa* (Philology Committee of the House of Rui Barbosa). It is certain that, for this, an extensive work was developed by Silva Neto, Nascentes, Cunha, and Rossi, masters and motivating forces of the dialectological studies as well as ideological precursors of the developments in several centers of studies in the country.

Besides giving dialectology courses in several colleges starting in 1943, Silva Neto also founded the Center of Studies of Brazilian Dialectology in the *Museu Nacional* in 1953. With the publication of *Guia para estudos dialetológicos* (1957), he proposes the creation of an annual course of Brazilian dialectology, aiming at guiding and coordinating research and elaborating on regional and national atlases through a single inquiry applied throughout the territories.

The importance of Nascentes, apart from the previously mentioned works, is also highlighted in the two volumes of *Bases para a elaboração do Atlas lingüístico do Brasil* (1958–1961), which are also indispensable manuals for dialectologists. In the first volume, the author recommends the creation of regional atlases. To guarantee the consistency of the work, he insists that we have to divide the country in regions, each one with its own superintendent; the superintendents would be under the direction of a technician, with a central supervising organ, The Philology Committee (Nascentes 1958, 8). He also develops a series of themes to design the questionnaire and—recognizing that the good determination of the inquiry points has great influence on the perfection of an atlas and that these points must be located in the middle of the most characteristic regions of the type of language—he presents a temporary proposal of 606 points of inquiry distributed throughout the Brazilian states and territories of the time. This proposal is continually revisited by researchers conducting any sort of dialectological research. In the second volume, Nascentes (1961) adds to the general questionnaire from the first volume, supplying precise details and examples of 100 phonetic phenomena and 116 questions of morphological and syntactic nature.

Celso Cunha, a great promoter of dialectology in Brazil, presented the proposal for preparing seven linguistic-ethnographic atlases of Brazil according to the divisions of Brazilian cultural areas defined by Manuel Diegues: *Amazônica, Nordeste Litoral, Nordeste Mediterrâneo, Planalto, Centro-Oeste* and *Centro-Leste e Extremo Sul* along with Silva Neto at the *III Colóquio Internacional de Estudos Luso-Brasileiros* in 1957.

Nelson Rossi, professor at the Universidade Federal da Bahia (UFBA) since 1955, created a laboratory of Experimental Phonetics at UFBA. His great accomplishments are the making of the *Atlas prévio dos falares baianos* and the creation of a team of researchers, among them Suzana Cardoso, President of the national committee of ALiB, and Jacyra Mota, Executive Director of ALiB.

1991 brings the pioneering work of Brandão, *A geografia linguística no Brasil*, an accessible read for the beginner and a safe guide for specialists.

The fourth phase starts in 1996, when researchers in the fields of dialectology, lexicology, lexicography, and sociolinguistics, guided by Cardoso and Mota, laid the foundations of the eagerly awaited *Linguistic Atlas of Brazil* (ALiB). In order to show the development of geolinguistic research in Brazil, Aguilera (1998; 2005) organized two works with finished and in-progress articles by organizers of atlases.

3 Geolinguistics in Brazil based on geographic regions and their respective centers of studies

Researchers usually deal with the history of geolinguistics in Brazil following the chronological order of the publications of each atlas. For this overview article, however, we chose to do it by geographic region, starting with the Northeast Region and Bahia.

3.1 Northeast Region: APFB, ALS, ALS II, ALPB, ALECE, ALiPE, ALiMA, and others of small domain

Researchers Dinah Callou, Carlota Ferreira, Jacyra Mota, Vera Rollemberg, Nadja Andrade and Suzana Cardoso were all born in Bahia. Rossi lived in the region, and the first state atlas came from there: the *Atlas prévio dos falares baianos* (APFB) (Rossi 1963), as well as the *Atlas linguístico de Sergipe* and the *Atlas linguístico de Sergipe II*. Additionally, the ALiB headquarters is at UFBA.

These three state atlases, in addition to the extreme scientific rigor in all stages of preparation, have in common the supervision by Nelson Rossi, UFBA as the project headquarters, their teams of investigators and transcribers, and the methodological orientations. In some aspects, they present differences, as synthesized in Table 2:

Table 2: Similarities and differences between the APFB, the ALS and AS. II.

Atlas	APFB	ALS	ALS II
Number of points	50	15	same as ALS
Informants per point	1–6	2	same
Age	25–84	30–65	same
Gender	M and / or F	F (A) and M (B)	same
Total of informants	100	30	same
Number of questions asked	182	674	same
Formulation of the questions	not controlled	same for all	same
Recorder	no	in reel-to-reel tape	same
Transcription	simultaneous	after recording	same
Dimensions	one-dimensional	two-dimensional	same
Charted questions	from the extract in one questionnaire	same as APFB	the inedited ones
Explanatory notes in the map	on the obverse	on the back page	same
Type of map	cartogram	cartogram	maps
Frequency charts	no	no	yes

The territorial extension of Bahia demanded a wider network of data collection points than that of Sergipe, the smallest Brazilian state. Rossi explains that, in order to decide on the localities in Bahia, they took Nascentes' proposal (1958) as a starting point—that is, to research the 39 localities. Considering, though, that the convenient data collection points to give a view of Bahia within the Brazilian extension could not be the most indicated to give a view of the facts pertaining to Bahia (1965, 21–22). They reviewed Nascentes' indications, cataloguing 1076 localities, from which they selected the 50 most representative ones, distributed throughout the 16 physiographic regions of the state. The team of investigators was formed by newly undergraduate language students, who were trained by Rossi. 182 questions, mostly of lexical and semantic nature, were extracted from the questionnaire, which initially comprised over 3000 questions. The answers to these questions were also used for the phonetic-phonological maps. The questions were previously tested in four localities. The themes of the questions concern nature and humans, emphasizing life in the countryside.

Of the 198 linguistic maps, 154 are analytical; that is, they present a transcription of the variant at each point with identification[7] of the informant(s). This allows the study of the phonetic phenomena, such as medium pretonic vowels, /r/ and /s/ in internal and external syllable coda, the pronunciation of /ti/ and /di/ and the proparoxytone, among others. Map 1, for example, presents the variants for *moon* and shows that the standard form *lua* is predominant in most points, but the historical nasalized forms *luma*, *lũma*, and *lũa* were still frequent among informants' speech in 13 of 50 localities, either as the only form or competing with the most modern variant. The other 44 maps summarize all previous maps and are mixed—the variants are represented by figures in the interior of the map and oriented by map keys placed at the bottom corner. This allows for a clearer and faster visualization of the distribution of the various subdialects.

The data in both atlases of Sergipe, ALS (Ferreira et al. 1987) and ALS II (Cardoso 2005), were collected through recordings on reel-to-reel tape by the same APFB team with a more systematic methodology: there were two[8] informants per locality, one of each sex, with an age range of 30 to 65. Additionally, the localities in APFB are numbered 1–50, with those from ALS and ALS II numbered 51–65, indicating the spreading of the *baiano* variety towards the state of Sergipe.

For ALS II and as part of his doctoral thesis, Cardoso reviewed the material collected by ALS and selected 161 unpublished questions on the area related to humans, for it contained in the set of questions which characterized it, two sub areas that are identified as follows: physical man—anatomy, illnesses, physical impairment—and social man—moral qualities, clothing, social activities (Cardoso 2005). As to its presentation, the ALS II is more innovative, featuring colored maps and the indication of waterways, roads, and railways as well as scale and frequency charts.

Dialectology in this region also presents a large production of monographic studies about regional economic activities, described through the lens of Lexicology, Lexicography, and Terminology, such as, for instance, fishing, fishermen, artisanal shipbuilding, horses, herdsmen and pineapple.

Aragão and Menezes (1984) published the *Atlas linguístico da Paraíba*, in two volumes. Aragão elaborated the project under the guidance of Manuel Alvar in Madrid and officially initiated it in 1974. During the years that followed, he devoted himself to preparing the team of investigators in both methods and techniques and beginning data collection in 1979 in 25 base localities. Each locality had three satellite municipalities, totaling 100 points and covering the whole state. The data registered on ALPB show a unique distribution of lexical and phonetic variants, pointing away from the existence of isoglosses.

---

7 In APFB, the letter A, placed beside the variant, indicates the informant whose profile is closer to the ideal; in ALS and ALS II, women are identified by the letter A and men by the letter B.

8 ALS is the first two-dimensional atlas published in Brazil. The second is ALPR.

The state atlases of this region, subsequent to ALPB, are no longer exclusively diatopic, and they incorporate the principles of Labovian sociolinguistics to traditional geolinguistics—geo-social linguistics or multi-dimensional geolinguistics was born. The *Atlas linguístico do Ceará* (ALECE) (Bessa et al. 2010) was released at the end of the 1970's and investigated 67 localities distributed throughout the state, consulting 265 male and female informants, ranging from 30 to 60 years old, whose education varied from none to having completed secondary school. Published in two volumes, the ALECE consists of 240 maps, of which 108 are lexical and 132 phonetic. It also contains a vocabulary with 515 entries.

The other works on multi-dimensional geolinguistics in the Northeast region are theses and essays, which originated the atlases by Pereira (2007), Sá (2013), Lima (2009), Almeida (2009), Monteiro (2011), and Silva (2012).

The *Atlas Linguístico de Pernambuco* (Sá 2013) is based on the ALiB methodology, and its phonetic data indicates the predominance of [di] and [ti] in words like *prateleira* 'shelf', *liquidificador* 'blender', *elefante* 'elephant', *tarde* 'afternoon', *tio* 'uncle'; of [s] in internal or final coda, like in *fósforo* 'match', and *três* 'three' in the rural localities, but alternating with [ʃ] only in the capital and surrounding localities; the [h] in internal coda is categorical; the medium pretonic vowels are open only in *borboleta* 'butterfly' and *elefante,* but they are kept or accentuated in *tomate*/*tumati* 'tomato', *gordura*/*gurdura* 'fat'. Lexical maps indicate that the rural forms are frequent in the countryside but do compete with the more urbanized versions such as *pinguela* 'footbridge' and *passagem* 'passage', *maré* 'tide', *mareta* and *onda* 'wave', *chuva de pedra, chuva de gelo e chuva de granizo* 'hail', *mocotó* and *tornozelo* 'ankle'.

The *Atlas Linguístico do Maranhão* (ALIMA) project, coordinated by Ramos since 2000, is in the process of mapping of data. The methodology follows ALiB orientation, with the addition of questions about specificities *maranhenses* 'from Maranhão' to the questionnaire, like *bumba-meu-boi*, *culinária* 'culinary', *cultural manifestations of African roots in Maranhão, products from agroextractivism* and *reggae*.

Doiron's thesis, the *Atlas linguístico e etnográfico de Alagoas* (ALEAL), is in progress at *Universidade Estadual de Londrina* (UEL). The network of points comprises 23 localities. In each one, two informants were surveyed—one male and one female between 30 and 50 years old with only minimal education. Of the nine states that form the Northeast Region, only Piauí does not have a geolinguistics project.

In the scope of geolinguistics and geo-sociolinguistics, over a dozen atlases have already been produced or are in the elaboration process, of which, two are one-dimensional (*), three are two-dimensional (**), and the others are multi-dimensional. Table 3 presents a synthesis of the data in these atlases.

Table 3: Geolinguistic and/or geo-social linguistic studies about the dialects in the Northeast Region.

Atlas	Author	Conclusion/ publication	Nature	University
APFB (*)	Rossi et al.	1963	institutional research	UFBA
ALPB (*)	Aragão/Menezes	1984	institutional research	UFPB
ALS (**)	Ferreira et al.	1987	institutional research	UFBA
ALS II (**)	Cardoso	2002/2005	thesis	UFBA
AGLiP	Pereira	2007	thesis	UFRJ
ALMASPE	Almeida	2009	essay	UFPB
ALIg-CE	Lima	2009	essay	UFC
ALCE	Bessa et al.	2010	institutional research	UFC
ALLSCA	Monteiro	2011	essay	UECE
ALiCOP	Silva	2012	essay	UFC
ALiPE	Sá	2013	thesis	UFPB
ALiMA	Ramos et al.	in progress	institutional research	UFMA
ALPLN-BA	Pereira	in progress	thesis	UFBA
ALEAL (**)	Silva-Doiron	in progress	thesis	UEL

3.2 Southeast Region: EALMG, ALES and others of small domain

The Southeast region is formed by Minas Gerais, Espírito Santo, Rio de Janeiro, and São Paulo. Even though the most important linguistic research centers in the country and the best postgraduate programs, as evaluated by CAPES, are concentrated here, it is not the center of geolinguistic studies. We believe that while dialectology advanced in the other regions, sociolinguistics installed itself and kept expanding, especially in Rio de Janeiro, attracting the new generations more than dialectology. However, the aforementioned renowned researchers in the field of dialectology are from the Southeast Region: Amaral, Nascentes, Silva Neto, Cunha, Teixeira, and, later, Caruso and Brandão, in addition to Callou and Santos in the field of geo-sociolinguistics.

The EALMG project (Ribeiro et al. 1977) was born out of the ideas of a group of recent undergraduates from Juiz de Fora College. José Dionísio Ladeira[9] provided the following testimony regarding the genesis of EALMG:

9 Received by e-mail, on the 13 may, 2015, from Dr. José Dionísio Ladeira, one of the interviewers for EALMG.

> 'I don't know exactly how the idea arose. But I know that Zágari and Passini—all of us at the time had only Undergraduate degree—took an intensive course in Florianópolis in summer of 1971 (or 72 or 73—I'm not sure) and Nelson Rossi, of *Atlas baiano*, was among the teachers. I think it came from there. [...] Mário Roberto Lobuglio Zagari, Antônio Pereira Gaio, and José Ribeiro were all 'incorporated' teachers within the college. José Passini started to participate, and shortly after that, he was hired directly by UFJF (1970). All were teachers.'[10]

The EALMG investigated 116 localities, interviewing male informants between 30 and 50 years old with little education. They were given a questionnaire with 415 questions. Indirect research in 302 localities complemented previous research. Projected for four volumes, only the first was published in 1977, featuring 73 linguistic maps, of which 45 are onomasiological, of lexical and phonetic-lexical character. 28 are isophonic and isolexic, covering only two semantic fields: time and child's play. The EALMG's data reveal the existence of three dialects in the *mineiro* territory (Minas Gerais), characterized by phonetic and also lexical aspects: i) the *paulista* variety from the South and from *Triângulo Mineiro*; ii) the *baiano* variety from the North; and iii) *mineiro,* which covers the region formed by the Metallurgic Zone, Mata and Vertentes.

Rio de Janeiro was, in the 1950's, the biggest center of dialectological studies with the House of Rui Barbosa, the *Centro de Estudos de Dialectologia Brasileira* while also being the cradle of Cunha and Silva Neto's proposal for elaborating the Linguistic-Ethnographic Atlas of Brazil by region. The *Atlas linguístico dos pescadores do Estado do Rio de Janeiro* (APERJ) initially coordinated by Cunha and, at the end, by Sílvia Brandão, was also conceived there. It is a project that subsidized important works which associated the principles of dialectology to those of sociolinguistics, like the thesis *Micro Atlas Fonético do Estado do Rio de Janeiro* (Micro AFERJ) (Almeida 2008) and the essay *Atlas do Entorno da Baía da Guanabara* (AFeBG) (Lima 2006).

The *Atlas Linguístico do Espírito Santo* (ALES), started by Rodrigues in 2006, concluded the collection of data in 35 rural points, where 35 men and 35 women between 30 and 55 years old with little education were interviewed. In 20 urban points, there were eight informants per locality, four of whom had college education. The project is in the data revision phase and is beginning the elaboration of the maps.

São Paulo doesn't have its own linguistic atlas, despite the pioneering work by Caruso from 1980 to 2000. Thanks to his effort, the dialectological mentality at Universidade Estadual Paulista (UNESP) in Assis was created, guiding dialectal monographs about handmade horticulture, fishing, cane syrup, and beekeeping, among others.

---

10 "Exatinho como surgiu a ideia não tenho conhecimento. Mas sei que Zágari e Passini—todos nós, à época, tínhamos apenas a Graduação—fizeram, em Florianópolis, um curso intensivo num dos verões do mil novecentos e setenta e um (ou dois, ou três, sei lá) e tiveram, entre os professores Nelson Rossi, do Atlas baiano. Acho que a ideia veio daí. [...] Mário Roberto Lobuglio Zagari, Antônio Pereira Gaio e José Ribeiro, todos professores 'incorporados' junto com a Faculdade. Um pouco depois começou a participar também José Passini, contratado já direto pela UFJF (1970). Todos professores."

Caruso (1983) published the questionnaire for the Linguistic Atlas of the State of São Paulo, which was tested in many *paulista* localities by public school teachers.

As Caruso said about the previous inquiry in the countryside of São Paulo: 'The result more than exceeded what we were looking for and, more than that, it showed that linguistic research about the variety in the state of São Paulo is most urgent and necessary for the dialectological knowledge of Brazil' (Carurso 1982, 69–70).[11]

Later, he initiated ALESP in 100 localities. A large number of the interviews were transcribed by Vandersí Castro and Harumi Pisciotta, graduate students at Universidade Estadual de Campinas (UNICAMP), led, at the time by Brian Head, who was responsible for ALESP continuity. With Brian Head's return overseas and Caruso's passing, the data was lost and, with that, a great part of the history of geolinguistics of both São Paulo and Brazil in general was gone.

Caruso (1982) presents four experimental maps with the data collected by mail in 65 localities: map 10: *fruto que nasce pegado ao outro* 'a fruit that has naturally grown together with another one'; map 24: *terçol* 'sty'; map 44: *estilingue* 'slingshot' and map 47: *bolinha de gude* 'marble'. Like Rossi, Caruso's great merit, in addition to form the ALESP Project, was having transmitted his dedication to dialectology to his students, leading them to form centers of dialectal studies like the ones at UFMS and UEL at their universities of origin.

At UNICAMP, Castro (2006) defended the thesis *A resistência de traços do dialeto caipira*, a study based on regional Brazilian linguistic atlases. In his research, he created 26 maps with data from EALMG and ALPR. Santos coordinates the research group of dialectology and geolinguistics (GPDG/USP) at USP and has oriented works of social geolinguistic nature about *paulista* lexical variation and atlases of other regions like Marajó by Cardoso da Silva (2002) and Goiás by Augusto (2012). At UEL, Santos-Ikeuchi (2014), in the essay *Atlas linguístico topodinâmico do Oeste de São Paulo* (ALTOSP), compares the data of the field research in seven localities from West São Paulo to the data found in atlases published in the Northeast region to verify the permanence or abandonment of dialect traces of migrants among first generation *paulista* descendants of the Northeastern region. In the Southeast region, as we saw, only one atlas has been published. Another is in progress, but both Rio de Janeiro and São Paulo have theses and essays which include atlases of small domain in studies of geo- and sociolinguistic nature.

In Table 4, we present a synthesis of the studies conducted on the dialects of the Southeast region.

11 "O resultado superou em muito aquilo que buscávamos e, mais do que isso, mostrou que a pesquisa linguística do falar do Estado de São Paulo é urgentíssima e necessária para o conhecimento dialetológico do Brasil."

Table 4: Geolinguistic and/or social dialectology studies about the dialects of the Southeast Region.

Atlas/Region	Author	Conclusion/ publication	Nature	University
EALMG	Ribeiro et al.	1977	institutional research	UFJF
Sorocaba	Santos	2002	essay	USP
Litoral Sul Paulista	Imaguire	2004	thesis	USP
AFeBG	Lima	2006	essay	UFRJ
Ilhabela	Encarnação	2006	essay	USP
Região do Grande ABC	Cristianini	2007	thesis	USP
MicroAFERJ	Almeida	2008	thesis	UFRJ
Iguape	Silveira	2009	essay	USP
Caraguatatuba, Ilhabela, São Sebastião e Ubatuba	Encarnação	2010	thesis	USP
Região Norte do Alto Tietê	Silva	2012	thesis	USP
ALTOSP	Santos-Ikeuchi	2014	essay	UEL
ALES	Rodrigues	in progress	institutional research	UFES

3.3 South Region: ALERS, ALPR, ALPR II, and others of small domain

In the South region, geolinguistics started at UFRGS with Bunse who, in the 1950's, recognized the important roles of Silva Neto, Nascentes, and Rossi as well as the First Brazilian Congress of Dialectology and Ethnography, held in Porto Alegre in 1958, in awakening interest in dialectological studies. Bunse (1969) reveals that, since 1851, some regional vocabularies and glossaries had started to be published in that state but just a few, though, are studies about the *gaúcho* dialect.

To attain knowledge about the southern reality and fruition of this atlas, Bunse (1969) authored monographs published in 1955 and 1960, about the lexicon of the coal mines, sugar cane, fishing and fisherman, mills, and mate herb. He also performed an indirect study, the results of which comprise nine succinct maps. Map 1 presents the distribution of the variants for *charrete* 'carriage', Map 2 *cerca de arame* 'wire fence', Map 4 for *farm* (fazenda), and Map 6 for *corvo* (urubu) (vulture). Finally, he initiated the collection of data in direct research, though the project has not been concluded.

The *Atlas linguístico-etnográfico da Região Sul* (ALERS) project (Koch et al. 2002; Altenhofen et al. 2011) started in 1980 at UFRGS with the purpose of continuing the re-

search initiated by Bunse. In 1987, a team of researchers from the three federal universities (Paraná, Florianópolis, and Rio Grande do Sul) was assembled. It established 275 rural localities (100 in Paraná, 95 in Rio Grande do Sul, and 80 in Santa Catarina) and 19 in the urban center. This is the first Brazilian linguistic atlas to apply questionnaires of an assorted nature (about 1000 questions): phonetic-phonological, lexical semantic and morphosyntactic. The questionnaire is also the first to cover all the states of a geographic region. The first volume (2002) consists of the introduction; the second contains phonetic and morphosyntactic maps. The volume with lexical semantic maps was published in 2011.

Furlan (1989) wrote a detailed description of the Azorean influence on the Portuguese of Santa Catarina, stressing characteristic and common traces as well as documenting the influence of other dialects. Two atlases of small domain, in addition to the ALERS maps, exist regarding the *catarinense* dialects: Imaguire's thesis (1999), defended at USP, about the dialects of the island of Santa Catarina and Guimarães's thesis (2008), about the dialects of São Francisco do Sul, defended at UEL.

The studies about the dialects of Paraná (*paranaenses*) follow the same trajectory: vocabularies and glossaries first, monographs and atlases later. The first records, it seems, are from Saint-Hilaire's notes, published in *Viagem à Comarca de Curitiba* (1820). He states that at the Campos Gerais, the majority of the population was formed by white men, that is, not *mestizos*, who, despite the little education, would express themselves better than individuals encountered in São Paulo, because they didn't give the *ch* the sound of *ts* and to the *g* the sound of *dz*, changes introduced into the Portuguese language by the Indians, with which the settlers of the districts of Castro and Curitiba didn't mingle much (Saint-Hilaire 1964, 18).

Leão (1994) published *Contribuições históricas e geographicas para o dicionário do Paraná*. Although not systematic, it is a valuable record as it offers data concerning the history of the *paranaenses* dialects. Besides Saint-Hilaire, others have also recorded phonetic and lexical details: Coelho[12] (1995) and Muricy[13] (1975).

The monographic studies centralized at the *Universidade Federal* and later at the *Faculdade de Letras de Ponta Grossa* involve field research done in the south of the state. Among these studies are the monographs by Amaral (1956) about the seaside municipality of Guaraqueçaba, Sêcco (1956) about the city of Palmeira, Toniolo (1967) about the language of the Gaúcha Central Depression and the Campos Gerais then later (1981; 2002) about the rural language of Tibagi; Mercer (1977; 1979), and Alvar and Alvar (1979) about Guaraqueçaba.

12 The first *paranaense* to publish a book, Salvador Coelho, a court judge, provincial representative, and writer, covered the path between the *paulista* capital and the city of Lapain (1860). In this travel report, he includes a vocabulary with 67 entries.

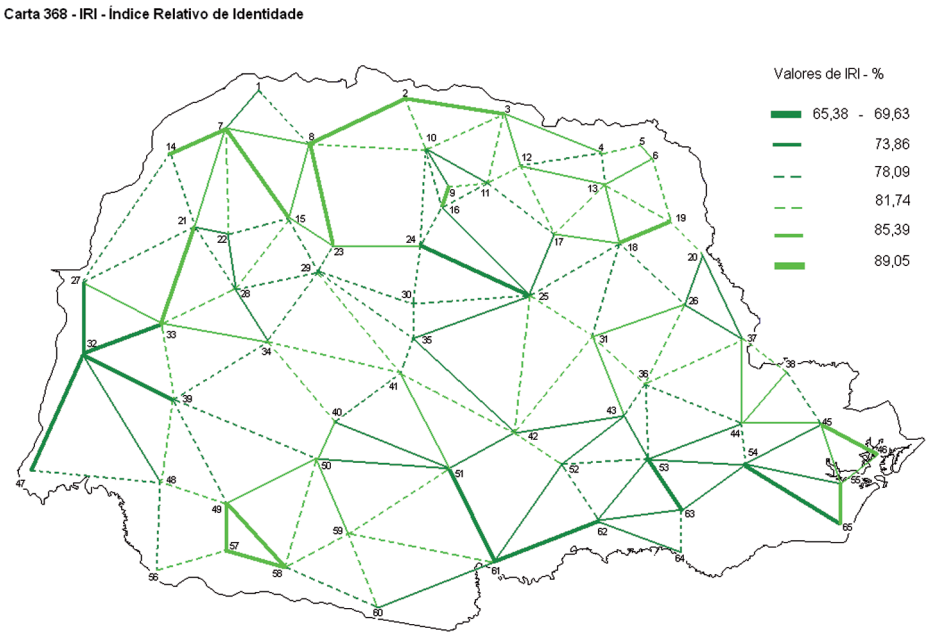
13 The General José Candido da Silva Muricy chronicled a trip to the then still unexplored countryside of Paraná during the last years of the 19th century. The work has a vocabulary of 278 words collected from the dialect of the inhabitants along the way.

França (1985) was the first scholar to describe the dialect of north Paraná, specifically the traces of the *caipira* dialect. The first references to Paraná's need of its own linguistic atlas, however, are conveyed by Wouk (1976) and Alvar and Alvar (1979). These were later materialized by Aguilera (1994).

Between 1985 and 1990, UEL started to establish itself as the second major center of dialectological studies in Paraná, influenced by the researchers Caruso and Hoyos Andrade from UNESP-Assis. The publications from UEL include the essay *Aspectos linguísticos da fala londrinense: a sketch of a linguistic atlas of Londrina (EALLO)* (1987), and the thesis *Atlas linguístico do Paraná (ALPR)*[14] (1990), both by Aguilera, UNESP-Assis campus. The network of ALPR points includes 65 localities in which men and women of rural origin between 27 and 62 years of age with minimal education were interviewed. Aguilera adopted the ALESP questionnaire (Caruso 1983), adapted to the circumstances of Paraná, and developed 191 maps—92 lexical, 70 phonetic—and 29 isoglosses.

The answers which were not mapped by Aguilera (1994) were used as a topic for Altino's (2007) thesis *Atlas linguístico do Paraná II* (ALPR II). The thesis consists of two volumes: the first presenting the actual thesis and the second, which is made up of 175 maps, 125 of which are lexical and 50 phonetic, following ALPR numeration from 192 to 368. The ALPR II is the first atlas in Brazil to include dialectometric results. The author also was innovative in his presentation of the maps. Following the ALS II model, it includes on the obverse of each map, the scale and the hydrographic basins. Tables, charts of frequency, and explanatory notes can be found on the back page.

14 The topic was suggested by Dr. Pedro Caruso, and the thesis, which was defended in 1990 and published in 1994, was directed by Dr. Raphael Hoyos-Andrade.



Map 2: Map 368 of the *Atlas Linguístico do Paraná*. vol. II (Altino 2007), mapping relative indices of identity.

Busse (2010), in his thesis *Um estudo geossociolingüístico da fala do Oeste do Paraná*, presents 96 maps—57 lexical and 39 phonetic. Data was collected from nine localities in the region using a questionnaire based on ALiB, ALERS and ALPR. Busse was innovative compared to the other Atlases of Paraná, when he inserts in each map three graphs with the frequency of the variants in regards to diasexual, diagenerational, and diastratic dimensions. The author concludes that, despite being a multicultural region and touristic route, the west of Paraná preserves some traces of two distinct dialects connected to the history of the settlement: the north, for example, is characterized by the predominant use of *mexerica* 'tangerine', *camomila* 'chamomile', *caçula* 'youngest child', *estilingue* 'slingshot', *geleia* 'jam' as well as by the [x] in the attack position, [ɹ] in internal and external syllable coda, and elevation of final middle unstressed vowels. In the dialect of the South, *bergamota* 'tangerine', *maçanilha* 'little apple', *nenê* 'baby', *bodoque* 'slingshot', *chimia* 'jam', and [ɾ] and [r] in initial and final positions as well as preserving the final middle unstressed vowels are more frequent.

In the thesis *Estudo geossociolinguístico dos falares paulistas*, Romano (2015) analyzed, using the ALiB corpus, the speech of 472 informants from 118 localities corresponding to the area that Nascents calls the *sulista* subdialect. Romano elaborated 71 maps generated by the computational tool [ΣGVCLin]—a software which generates and visualizes linguistic maps. He concludes:

> 'from the observation of the diatopic behavior of the documented variants (*córrego, tangerina, menino, bolinha de gude,* and *geleia*) and the statistical treatment to validate hypotheses about the lexical homogeneity and/or heterogeneity in the area of the *sulista* subdialect, we conclude that the northern portion of the investigated territory presents differences in relation to the southern part, showing the existence of two major dialects: *paulista* and *sulista*, despite the borders and coverage of these two dialects being virtual and fluid' (Romano 2015, p. 7).

Romano presents three innovations to geolinguistics: the statistical treatment of data to validate hypotheses, the first set of computerized maps [ΣGVCLin], created by Seabra, Romano, and Oliveira for linguistic atlases, and, finally, the proof that Nascentes' (1953) *sulista* subdialect represents, in fact, two subdialects: *sulista* and *paulista*. Silva's thesis deals specifically with the distribution of the *caipira* or retroflex /r/ in the capitals and countryside of the Southeast region. In the synthesis of the works, Table 5 indicates one-dimensional (*) and two-dimensional (**) atlases; multi-dimensional works are without indication:

Table 5: Geolinguistic and/or geo-sociolinguistic studies about the dialects of the South Region.

Atlas	Author	Year	Nature	University
Aspectos [...] londrinense: esboço (*)	Aguilera	1987	essay	UNESP
ALPR (**)	Aguilera	1990/1994	thesis	UNESP
Estudo[...] da Ilha de Santa Catarina (**)	Imaguire	1999	thesis	USP
Aspectos [...] Cândido de Abreu: um estudo geossociolingüístico	Lino	2000	essay	UEL
Pelos [...] paranaense: [...] léxico popular de Adrianópolis	Altino	2001	essay	UEL
ALERS (*)	Koch et al.	2002	interinstitutional	UFRGS
ALPR II (**)	Altino	2007	thesis	UEL
Um [...] Oeste do Paraná	Busse	2010	thesis	UEL
ALERS: cartas léxico-semânticas (*)	Altenhofen et al.	2011	interinstitutional	UFRGS
AGELO	Romano	2012	essay	UEL
Em busca [...] áreas lexicais no Centro-Sul do Brasil	Romano	2015	thesis	UEL
Amaral (1920)[...]? /r/ caipira, no Sudeste brasileiro	Silva	in progress	thesis	UEL

3.4 Central-West Region: ALMS, ALiMAT, ALIGO, ALiTI, and others of small domain

According to our research, the Central-West Region (Goiás, Mato Grosso, and Mato Grosso do Sul), in contrast to other regions, largely lacked many dialectological studies prior to Teixeira's work (1944) in Goiás. Regarding Mato Grosso, research is quite recent, especially regarding the *cuiabano* dialect. Research on the dialect dates back to the 1970's with Drummond's work (1978) and Oliveira's essay about the language of the region's fishermen. Nogueira (1989), from UFMS, defended his thesis, *A linguagem do homem pantaneiro*, at UNESP-Assis campus. This research gap can be justified by and reflects the change happening in Brazil which started at the end of the 1950's with the construction of Brasília and the transfer of the capital of Brazil to the Central Plateau and, one decade later, with the creation of state and federal universities.

The region, however, has a considerable collection of atlases nowadays, all based on multi-dimensional geolinguistics, of which we highlight: (i) published: *Atlas linguístico do Mato Grosso do Sul* (ALMS) from Oliveira (2007); (ii) atlases concluded as theses or essays: Augusto (2012), Cuba (2009; 2015), Reis (2013), and Alencar (2013); (iii) atlases in progress: *do Mato Grosso* (ALiMAT) (Lima et al.) and *do Pantanal sul-mato-grossense*: a lexical study about the dialect of the *pantaneiro* man in real time and apparent time, by Alencar. In Table 6, we present a synthesis of the geo-social linguistic studies of the dialects in the Central-West Region:

Table 6: Geo-social linguistic studies about the dialects of the Central-West Region.

Atlas	Author	Year	Research	University
ALiPP	Reis	2008	essay	UFMS
ALMSEMT	Cuba	2009	essay	UFMS
ALMS	Oliveira	2007	institutional	UFMS
ALiGO	Augusto	2012	thesis	USP
ALCL	Alencar	2013	essay	UFMS
ALF-BRPY	Reis	2013	thesis	UEL
ALiTI	Cuba	2015	thesis	UEL
ALIMAT	Lima et al.	in progress	institutional	UFMT
Pantanal Sul-Mato-Grossense [...]	Alencar	in progress	thesis	UFMS

From Table 6, we highlight the ALMS (Oliveira 2007), the first linguistic atlas of the region. It was initiated by Nogueira and finalized by Oliveira and his team. The network has 32 points, in each of which four socially stratified informants participated. The questionnaire, featuring 557 questions, was based on Caruso (1982) with adaptations

made to reflect the reality of Mato Grosso do Sul. The results produced 55 phonetic maps—153 lexical and seven morphosyntactic.

Also worthy of note is the *Atlas linguístico topodinâmico do território incaracterístico* (ALiTI) (Cuba 2015) due to the fact that it covers seven localities in Mato Grosso, two in Rondônia, one in Goiás, and one in Tocantins, corresponding approximately to the territory defined by Nascentes in his dialectal division of Brazil. Cuba interviewed four informants per locality, two of whom, in age group II (50–65), were not born in the locality. In the end, the author constructed a lexical synthesized map and another phonetic synthesized map, concluding:

> 'The data points to a tendency towards a southern kind of dialect (*paranaense, catarinense,*and *rio-grandense*) in the central part of the area investigated and near the roads BR 163 and BR 364, while in localities farther away, the linguistic variety tends towards the Amazonian varieties [...] In sum, we infer that to the shadows of history which permeated this geographic area for over a century, this territory is no longer uncharacteristic, but preserves within itself the multifaceted properties of a linguistic reality that can no longer be ignored' (Cuba 2015, 228–229).[15]

3.5 North Region: ALiSPA, ALAM, ALIPA, ALiRO, ALAP, ALITTETO and others of small domain

Dialectological works concerning the dialects of the north region before the 21th century are scarce (Veríssimo 1884; Miranda 1968). Starting in 2000, there has been an increase in interest in the *amazonense* dialects in the region, especially due to the ALiB project, released at the end of 1996 as well as to the leadership of professors Razky (UFPA) and Cruz-Cardoso (UFAM).

In Table 7, we present a list of the main geo-social linguistic works on the dialects of the North.

Table 7: Geo-social linguistics studies about the dialects of the North Region.

Atlas	Author	Conclusion/publication	University
ALISPA	Razky	2004	UFPA
ALAM	Cruz	2002	UFRJ

15 "os dados apontam para a tendência de um falar mais sulista (paranaense, catarinense e rio-grandense), na parte central do espaço pesquisado e nas proximidades das rodovias BR 163 e BR 364, enquanto em localidades mais afastadas, a variedade linguística tende à amazônica [...]. Em suma, inferimos que à sombra da história que permeou esse espaço geográfico por mais de um século, esse território já não é incaracterístico, mas guarda em si as propriedades multifacetadas de uma realidade linguística que não pode mais ser ignorada."

Table 7: (continued)

Atlas	Author	Conclusion/publication	University
ALIPA	Razky	in progress	UFPA
ALiRO	Teles	in progress	UFRO
ALAP	Razky et al.	in progress	UFPA/UFAP
ALITETTO	Silva-Poreli	in progress	UEL
ALSAM	Maia	in progress	UEL
AFBAM	Brito	2010	UFAM
ALFARIN	Justiniano	2010	UFAM
Baixo Amazonas (PA) and Médio Solimões (AM)	Azevedo	2013	UFSC/UFAM

The *Atlas linguístico sonoro do Pará* (ALiSPA) by Razky (2004) is the first of the region and the first sound atlas of Brazil. ALiSPA allows an automatic visualization of 636 phonetic maps from two observations: (i) from the context where the phoneme occurs and (ii) only of the variation of the phoneme and an automatic calculation of the number of variants. The *Atlas linguístico do Amazonas* (ALAM), presented in Cruz's thesis (2002) defended at UFRJ, used a phonetic questionnaire with 156 questions and a lexical-semantic questionnaire with 327 questions, which generated 107 phonetic and 150 lexical-semantic maps. Atlases of geo-sociolinguistic nature and based on the methodology of the ALiB are being developed in Pará (AliPA), Amapá (ALAP), and Rondônia (ALiRO).

The *Atlas linguístico topoestático e topodinâmico de Tocantins* (ALiTTeTO), in searching for the diasexual, diagenerational, diatopic-topostatic and diatopic-cinetic differences, proposes the charting of the aspects of the *tocantinense* dialect, which would make possible the verification of the influences that speakers from other Brazilian regions have had since the migration process from 1980 until present.

4 The Linguistic Atlas of Brazil and Nascentes's dialectal division (1953)

The *Atlas linguístico do Brasil* (ALiB) project,[16] approved during the *Caminhos e perspectivas para a Geolinguística no Brasil* seminar in November 1996 in Salvador, constituted a National Committee to coordinate it at the time. Initially, the participants

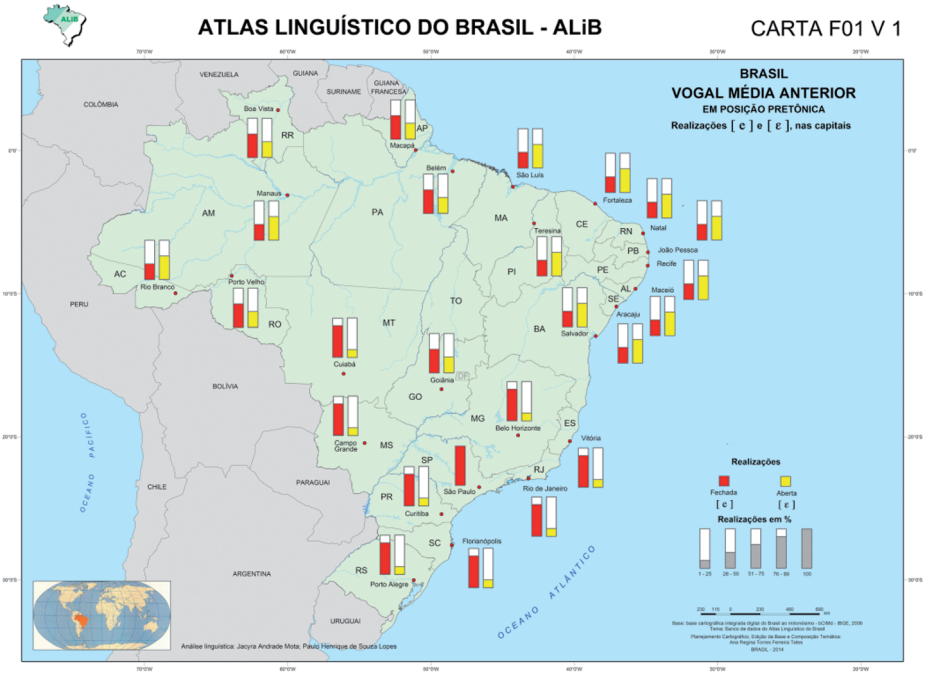
16 More detailed information is available at: www.alib.ufba.br, last accessed 18.02.2022.

were the authors of the finished state atlases and one of an atlas still in development, representing the federal universities UFBA, UFPB, UFJF and UFRGS and the State University of Londrina (UEL).

The ALiB, recognizing the importance of aggregating other social variables to the geographic variable, structured itself within the theoretical-methodological principles of multi-dimensional geolinguistics or geo-social linguistics. It consists of a network of 250 points, distributed over 26 Brazilian states. In each point, four informants were interviewed, stratified by sex, age, and education—with a maximum of eight years of basic education. Four informants with higher education were included in the capitals. The questionnaires were elaborated upon in order to register phonetic, lexical, prosodic, morphosyntactic, pragmatic, and metalinguistic variations, while also including questions for a semi-directed discourse and a text to be read. The interviews were carried out by 30 inquirers between 2001 and 2013.

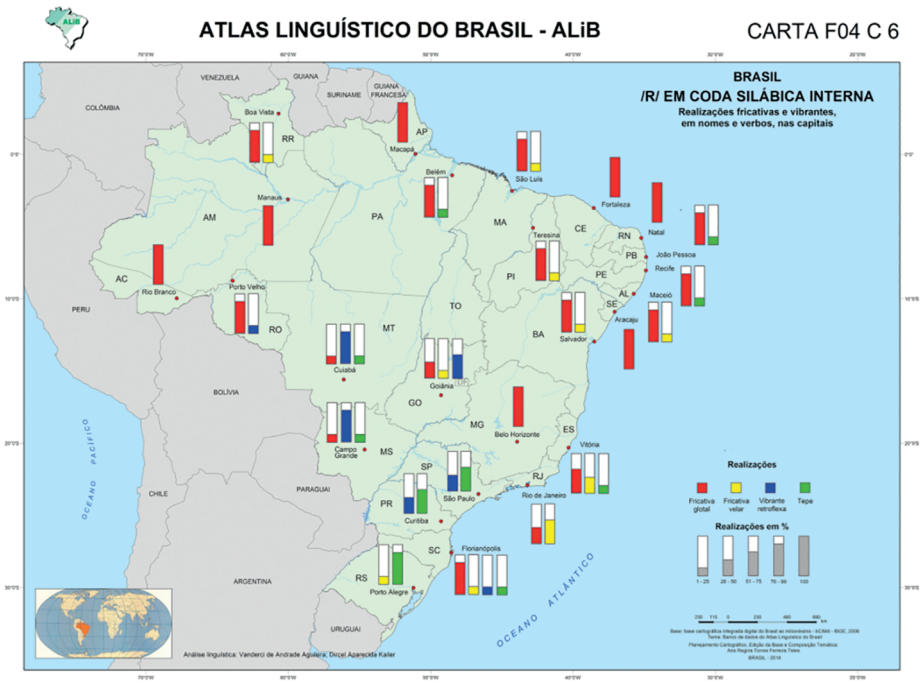
The ALiB will have several volumes and the first two ones were published in 2014. They contain the publication of the results from the 25 capitals: the first one features the introduction and the second one maps with phonetic, lexical, and morphosyntactic facts collected from 200 informants, eight in each one of the capitals.

The volume with the maps presents a diatopic and diastratic distribution of seven phonetic phenomena, from which we selected two for comments. The maps regarding the medium pretonic vowels, considered by Nascentes (1953) as a distinctive factor between the dialects of the North and South, show that the vowels /e/ and /o/ in the considered lexical items (i) remain closed in 100 % of the records in the capital São Paulo, (ii) predominate in the other capitals of the Southeast, South and Central-West, (iii) both variants occur in Goiânia, as well as in Belém, Macapá, Boa Vista and Porto Velho, and (iv), the open vowels dominate in the other capitals of the North and Northeast.



Map 3: Phonetic Map 01—vowels 1 (Cardoso et al. 2014).

Regarding the /r/ in internal syllable coda in the capitals of the North and Northeast, the glottal fricative is practically exclusive in Belo Horizonte (Southeast) and Florianópolis (South). The velar /ɾ/ has a high frequency in Vitória and Rio de Janeiro. The tap /ɾ/ predominates in São Paulo, Curitiba, and Porto Alegre, and the retroflex /ɾ/ prevails in Cuiabá, Campo Grande, and Goiânia, and has frequencies between 26 % in São Paulo and 50 % in Curitiba.

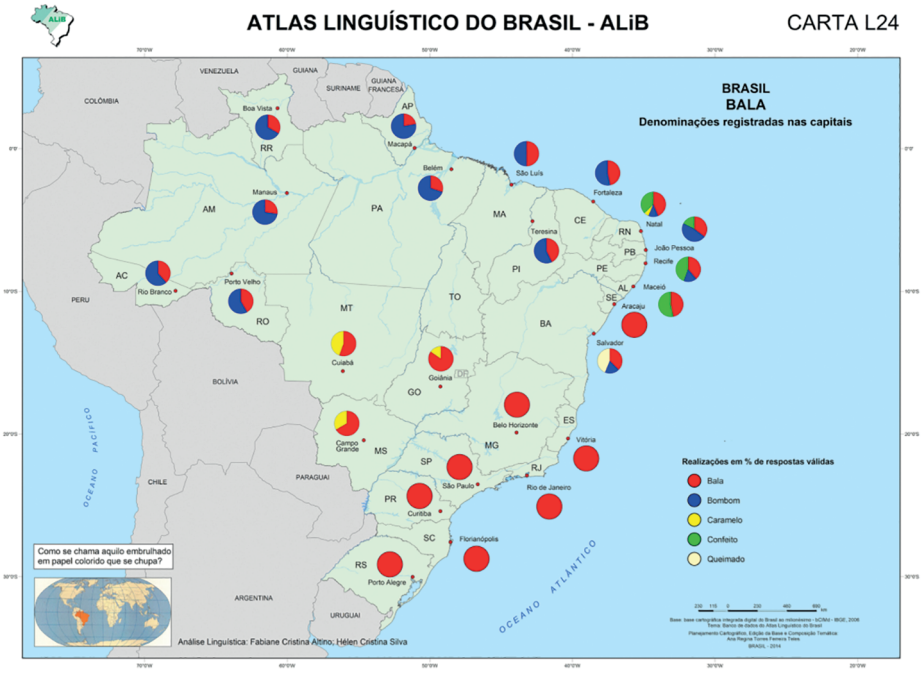


Map 4: Phonetic Map 04—consonants 6 (Cardoso et al. 2014).

The lexical maps refer to the distribution of the variants of 27 items, from which we selected two for comments. With regard to the *raiz branca comestível* 'edible white root', there is a well delimited diatopic distribution: in the North and Northeast, *macaxeira* is exclusive in all the capitals, except for Natal, where *aipim* was elicited by 25 % of the informants; in Fortaleza and Teresina, *mandioca* was recorded by the other 25 %. This variant is exclusive in Salvador, Porto Alegre, Florianópolis and is dominant in Vitória, Rio de Janeiro and Curitiba. In Curitiba, *mandioca* represents 45 % of the records and is exclusive in the dialect of the other capitals of the Southeast and Central-West.

Map 5: Lexical Map 08 (Cardoso et al. 2014).

As to the variants for *doce pequeno embrulhado em papel* 'little sweet wrapped in paper', in the capitals of the North and in São Luís, Teresina, Fortaleza and João Pessoa, the predominant form is *bombom*, although *bala* is present in the Northern and Northeastern dialects. The variant *confeito* is preferred in Recife and Maceió and was also recorded in Natal and João Pessoa. In Salvador, along with *bala* and *bombom, queimado* predominates. In the capitals of the Southeast and South, the only recorded form is *bala*, majoritarian in the capitals of the Central-West, where the variant *caramelo* is also frequent.



Map 6: Lexical Map 24 (Cardoso et al. 2014).

5 Conclusions

The dialects of Brazilian Portuguese—dialect being defined as regional variation—have been a reason for reflection and discussion by language scholars since the beginning of the 19th century, often fueled by separatist and nationalist ideals. Influenced by European dialectology, the first efforts to build a linguistic atlas of Brazil were started to clarify one major question: are the dialectal differences in Brazilian Portuguese of diatopic or diastratic nature? In this article, we reviewed the three stages or basal methods of dialectology, presenting, in a synthetic way, an overview of the evolution of this branch of Linguistics relative to the lexicographic, monographic, geolinguistic, and, nowadays, geo-social linguistic works. In this aspect, we verified that some states or regions had an earlier start or were more productive in the publication of dialectal studies over the last two centuries.

Regarding geolinguistics, specifically, we observed a considerable advance in several aspects:

i) Methodology: (a) Quantity and profile of the informants: by NORMS (non-educated, adult, rural worker, man and sedentary) a single informant, includes gender variable, different age groups, level of education, urbanized, fixed in the locality or

from other areas. In summary, a change from one-dimensional to two- and multidimensional and from topostatic to topodynamic atlases. (b) Instrument of data collection: In the beginning, there was a single questionnaire, which made the comparison of data difficult. ALiB questionnaires feature adaptations to local or regional specificities. (c) Means of data collection: originally implementing simultaneous transcription without the help of a voice recorder, the methodology has now evolved to use more sophisticated recording devices and specific software for more consistent transcriptions.

ii) Presentation of data: (a) from cartograms, that is, plain maps, without any concern for geopolitical data such as scale, boundaries, hydrography to more elaborated maps, containing these elements, additionally containing charts of frequency and informants' comments; (b) from handmade maps to maps programmed in special software, minimizing cartographic mistakes and greatly streamlining map printing; (c) from printed maps exclusively to audio and online maps.

iii) Study centers located exclusively in the Northeast, Southeast, and South: UFBA, UFRJ, UFJF, UFPB, UFPR and UEL; today more than 20 universities work in this field, with significant production in the North and Central-West.

In sum, we can conclude the following regarding the predominance of social over diatopic differences in BP: (a) there are, in fact, phonetic and lexical realities that mark dialectal areas, independent of social variables, but isogloss lines remain fluid and even tenuous; (b) there are other phonetic and morphosyntactic phenomena that are influenced by education, place of origin (rural or urban), age, and sex of the subject, mainly the less socially prestigious phenomena, such as the iotization and the rhoticism of /l/ in syllable coda or in consonant clusters; (c) the dialectal division proposed by Nascentes (1953) is still valid, at least in regard to the pretonic vowels, as shown in the map F01 V1 with data from the capitals; (d) other phonetic features, like the distribution of variants for the retroflex /ɾ/ in internal coda or the phonetic map F04 C6 present a singular distribution, in accordance with the socio-historical context of Brazil. As to the diatopic distribution of lexical variants, Maps L08 and L24 illustrate the division between the dialects from the North and South (Nascentes 1953) based on data from the capitals. Once all the data from the interior of Brazil have been mapped, we will have the conditions to safely evaluate the dialectal divisions of Brazil, to confirm or reject Nascentes' proposal (1953).

6 References

Aguilera, Vanderci de Andrade (1987), *Aspectos linguísticos da fala londrinense: esboço de um atlas linguístico de Londrina*, Curitiba, Conselho de Ciência e Tecnologia do Paraná.

Aguilera, Vanderci de Andrade (1994), *Atlas Linguístico do Paraná*, Curitiba, Imprensa Oficial.

Aguilera, Vanderci de Andrade (ed.) (1998), *A geolingüística no Brasil: caminhos e perspectivas*, London, EDUEL.

Aguilera, Vanderci de Andrade (ed.) (2005), *A geolingüística no Brasil: trilhas seguidas, caminhos a percorrer,* London, EDUEL.
Alencar, Beatriz Aparecida (2013), *Atlas Linguístico de Corumbá e Ladário: uma descrição da língua portuguesa falada no Extremo Oeste de Mato Grosso do Sul*, Master's thesis, Campo Grande, Universidade Federal do Mato Grosso do Sul.
Alencar, Beatriz Aparecida (in progress), *Atlas Linguístico do Pantanal sul-mato-grossense: um estudo lexical sobre o falar do homem pantaneiro em tempo real e em tempo aparente*, unpublished, Campo Grande, UFMS.
Almeida, Edilene (2009), *Atlas linguístico da Mata Sul de Pernambuco*, Doctoral thesis, João Pessoa, UFPB.
Almeida, Fabiana da Silva Campos (2008), *MicroAtlas Fonético do Estado do Rio de Janeiro* (MicroAFERJ), Doctoral thesis, Rio de Janeiro, UFRJ.
Altenhofen, Cleo Vilson/Klassmann, Mário Silfredo (edd.) (2011), *Atlas Linguístico-Etnográfico da Região Sul do Brasil (ALERS): cartas semântico-lexicais*, Porto Alegre/Florianópolis, EdUFRGS/EdUFSC.
Altino, Fabiane Cristina (2001), *Pelos caminhos da Geolinguística paranaense: um estudo do léxico popular de Adrianópolis*, Master's thesis, London, UEL.
Altino, Fabiane Cristina (2007), *Atlas Linguístico do Paraná II*, Doctoral thesis, London, UEL.
Alvar, Júlio/Alvar, Janine (1979), *Guaraqueçaba, mar e mato*, Curitiba, Universidade Federal do Paraná.
Amaral, Amadeu ([3]1976, [1]1920), *O dialeto caipira*, São Paulo, HUCITEC, Secretaria da Cultura, Ciência e Tecnologia.
Amaral, Serafina Traub Borges do (1956), *Contribuições para um inquéito linguítico do litoral do Paraná*, Letras 5–6, 157–166.
Aragão, Maria do Socorro Silva de/Menezes, Cleuza Bezerra de (1984), *Atlas linguístico da Paraíba*, 2 vol., Brasília, UFPB/CNPq.
Augusto, Vera Lúcia Dias dos Santos (2012), *Atlas semântico-lexical do Estado de Goiás*, Doctoral thesis, São Paulo, USP.
Azevedo, Orlando da Silva (2013), *Aspectos dialetais do Português da Região Norte do Brasil: um estudo sobre as vogais pretônicas e sobre o léxico do Baixo Amazonas (PA) e no Médio Solimões (AM)*, Master's thesis, Florianópolis, UFSC/UFAM.
Balbi, Adriano (1826), *Introduction à l'atlas ethnographique du globe*, Paris, Chez Rey et Gravier.
Bessa, José Rogério Fontenelle, et al. (2010), *Atlas linguístico do estado do Ceará*, 2 vol., Fortaleza, EdUFC.
Brandão, Silvia Figueiredo (1991), *A geografia lingüística no Brasil*, São Paulo, Ática.
Brito, Roseanny Melo (2010), *Atlas dos Falares do Baixo Amazonas* (AFBAM), Master's thesis, Manaus, UFAM.
Bunse, Heinrich (1969), *Estudos de Dialetologia no Rio Grande do Sul*, Porto Alegre, Edições da Faculdade de Filosofia, UFRGS.
Busse, Sanimar (2010), *Um estudo geossociolinguístico da fala do Oeste do Paraná*, Doctoral thesis, London, UEL.
Cardoso da Silva, M. P. S. (2002), *Estudo semântico-lexical com vistas ao Atlas Lingüístico da Mesorregião do Marajó/Pará*, Doctoral thesis, São Paulo, USP.
Cardoso, Suzana (2005), *Atlas lingüístico de Sergipe II*, Salvador, EDUFBA.
Cardoso, Suzana, et al. (2014), *Atlas Linguístico do Brasil*, 2 vol., London, EDUEL.
Caruso, Pedro (1982), *Amostra de um inquérito lingüístico prévio para o Estado de São Paulo*, ALFA 26, 69–78.
Caruso, Pedro (1983), *Atlas linguístico do Estado de São Paulo: questionário*, Assis, UNESP, Prefeitura Municipal de Assis.

Castilho, Ataliba Teixeira de (1972–1973), *Rumos da Dialetologia portuguesa*, ALFA 18/19, 115–153.
Castro, Vandersi Sant'Ana (2006), *A resistência de traços do dialeto caipira: estudo com base em atlas Lingüísticos regionais brasileiros*, Doctoral thesis, Campinas, UNICAMP.
Coelho, Salvador José (1995, [1]1860), *Passeio à minha terra*, Curitiba, Fundação Cultural.
Corrêa, Hydelvídia (1980), *O falar do caboco: aspectos fonético-fonológicos e léxico-semânticos de Itacoatiara e Silves*, Master's thesis, Rio de Janeiro, PUC.
Cristianini, Adriana Cristina (2007), *Atlas semântico-lexical da Região do Grande ABC*, Doctoral thesis, São Paulo, USP.
Cruz, Maria Luiza (2002), *Atlas linguístico do Amazonas*, Doctoral thesis, 2 vol., Rio de Janeiro, UFRJ.
Cuba, Marigilda Antônio (2009), A*tlas lingüístico da Mesorregião Sudeste de Mato Grosso*, Master's thesis, Campo Grande, UFMS.
Cuba, Marigilda Antônio (2015), *Atlas linguístico topodinâmico do Território Incaracterístico* (ALiTI), Doctoral thesis, 2 vol., London, UEL.
Doiron, Maranúbia Pereira Barbosa (2016), *Atlas linguístico-etnográfico de Alagoas* (ALEAL), Doctoral thesis, unpublished, London, UEL.
Encarnação, Márcia Regina (2006), *Estudo geolingüístico de aspectos semântico-lexicais nas comunidades tradicionais do município de Ilhabela*, Master's thesis, São Paulo, USP.
Encarnação, Márcia Regina (2010), *Atlas semântico-lexical de Caraguatatuba, Ilhabela, São Sebastião e Ubatuba*, Doctoral thesis, São Paulo, USP.
Ferreira, Carlota/Cardoso, Suzana (1994), *A Dialetologia no Brasil*, São Paulo, Contexto.
Ferreira, Carlota, et al. (1987), *Atlas linguístico de Sergipe*, Salvador, UFBA/Fundação Estadual de Cultura de Sergipe.
França, Tadeu (1985), *Marcas de fala caipira norte-paranaense. Bauru: Assembléia Legislativa do Estado do Paraná*, Master's thesis, Bauru, Universidade do Sagrado Coração.
Furlan, Oswaldo Antônio (1989), *Influência açoriana no Português do Brasil em Santa Catarin*a, Florianópolis, EdUFSC.
Guimarães, Tânia Braga (2008), *Atlas Lingüístico de São Francisco do Sul-SC*, Doctoral thesis, London, UEL.
Imaguire, Ligia Maria (1999), *Estudo com vistas a um Atlas linguístico da Ilha de Santa Catarina*, Master's thesis, São Paulo, USP.
Imaguire, Ligia Maria (2004), *Estudo geolingüístico de alguns municípios do Litoral Sul Paulista: abordagem de aspectos semântico-lexicais*, Doctoral thesis, São Paulo, USP.
Justiniano, Jeiviane dos Santos (2010), *Atlas Linguístico dos Falares do Alto Rio Negro*, Master's thesis, Manaus, UFAM.
Koch, Walter/Klassmann, Mário Silfredo/Altenhofen, Cleo Vilson (edd.) (2002), *Atlas Linguístico-Etnográfico da Região Sul do Brasil* (ALERS), 2 vol., Porto Alegre/ Florianópolis/Curitiba, EdUFRGS/EdUFSC /EdUFPR.
Leão, Ermelino Agustinho (1994), *Contribuições históricas e geographicas para o dicionário do Paraná,* 7 vol., Curitiba, Instituto Histórico, Geográfico e Etnográfico Paranaense.
Lima, Luciana Gomes de (2006), *Atlas fonético do entorno da Baía de Guanabara* (AFeBG), Master's thesis, Rio de Janeiro, UFRJ.
Lima, Fabiana dos Santos (2009), *Atlas linguístico do Iguatu-CE*, Master's thesis, Fortaleza, UFC.
Lino, Fádua Maria Moisés (2000), *Aspectos lingüísticos da fala de Cândido de Abreu: um estudo geossociolingüístico*, Master's thesis, London, UEL.
Marroquim, Mário (1996, [1]1934), *A Língua do Nordeste: Alagoas e Pernambuco*, Curitiba, HD Livros.
Mercer, José Luiz da Veiga (1977), *Notas dialetológicas sobre Guaraqueçaba*, Estudos Brasileiros 3, 35–64.
Mercer, José Luiz da Veiga (1979), *Le lexique des pêcheurs de Guaraqueçaba (Brésil)*, Doctoral thesis, 3 vol., Toulouse, Université de Toulouse II.

Miranda, Vicente Chermont de (1968), *Glossário Paraense*, Belém, Universidade Federal do Pará.
Monteiro, Jamile dos Santos (2011), *Atlas Linguístico Léxico-Semântico de Capistrano-CE*, Master's thesis, Ceará, UFCE.
Muricy, José Candido da Silva (1975, [1]1896), *Viagem ao paí dos jesuítas*, Curitiba, Imprensa Oficial do Paraná.
Nascentes, Antenor (1952), *Études dialectologiques au Brésil*, Orbis I/1, 181–184.
Nascentes, Antenor (1953, [1]1923), *O linguajar carioca*, Rio de Janeiro, Simões.
Nascentes, Antenor (1958–1961), *Bases para a elaboração do atlas lingüístico do Brasil*, 2 vol., Rio de Janeiro, Casa de Rui Barbosa.
Nogueira, Albana Xavier (1989), *A linguagem do homem pantaneiro*, Doctoral thesis, Assis, UNESP.
Oliveira, Alzira de (1980), *A linguagem dos pescadores de Mato Grosso: um estudo lingüístico-etnográfico*, Master's thesis, Rio de Janeiro, PUC.
Oliveira, Dercir (ed.) (2007), *Atlas Linguístico do Mato Grosso do Sul* (ALMS), Campo Grande, Editora UFMS.
Pereira, Maria das Neves (2007), *Atlas geolinguístico do Litoral Potiguar*, Doctoral thesis, Rio de Janeiro, UFRJ.
Pereira, Thais Dultra (2011), *A linguagem da pesca em comunidades baianas*, Master's thesis, Salvador, UFBA.
Razky, Abdelhak (2004), *Atlas Lingüístico sonoro do Estado do Pará* (ALiSPA), Belém, s/ed., (CD-ROM).
Reis, Regina Coelho Pereira (2013), *Variação linguística do Português em contato com o Espanhol e o Guarani na perspectiva do Atlas Linguístico-Contatual da fronteira Brasil-Paraguai* (ALF-BRPY), Doctoral thesis, London, UEL.
Ribeiro, José, et al. (1977), *Esboço de um Atlas Lingüístico de Minas Gerais*, Rio de Janeiro, Fundação Casa de Rui Barbosa.
Romano, Valter Pereira (2011), *Atlas Linguístico de Londrina: um estudo geossociolinguístico*, 2 vol., Master's thesis, London, UEL.
Romano, Valter Pereira (2015), *Estudo geossociolinguístico dos falares paulistas*, 2 vol., Doctoral thesis, London, UEL.
Rossi, Nelson (1963), *Atlas prévio dos falares baianos*, Rio de Janeiro, Instituto Nacional do Livro.
Rossi, Nelson (1965), *Atlas prévio dos falares baianos: introdução, questionário comentado, elenco das respostas transcritas*, Rio de Janeiro, Instituto Nacional do Livro.
Sá, Edmilson José de (2013), *Atlas linguístico de Pernambuco*, Docotral thesis, João Pessoa, UFPB.
Saint-Hilaire, Auguste (1964), *Viagem à Comarca de Curitiba (1820),* São Paulo, Nacional.
Santos, Sônia (2002), *Estudo geolingüístico de aspectos semântico-lexicais do campo semântico "Alimentação e cozinha" (Questionário do ALiB) no município de Sorocaba*, Master's thesis, São Paulo, USP.
Santos-Ikeuchi, Ariane (2014), *Atlas linguístico topodinâmico do Oeste de São Paulo* (ALTOSP), Matser's thesis, London, UEL.
Sêcco, Glacy Camarga (1956), *Aspectos lingüísticos palmeirenses: pesquisa dialectal*, Uniletras 5, 7–31.
Silva Neto, Serafim (1957), *Guia para estudos dialetológicos*, 2ª ed., Belém: Conselho Nacional de Pesquisa, Instituto Nacional de Pesquisas da Amazônia.
Silva, Moises Batista da (2012), *Atlas linguístico do Centro Oeste Potiguar*, Doctoral thesis, Fortaleza, UFC.
Silva, Rita de Cássia da (2012), *Atlas semântico-lexical da Região Norte do Alto Tietê-SP*, Doctoral thesis, São Paulo, USP.
Silveira, Rita de Cássia da (2009), *Estudo sociogeolinguístico do município de Iguape: aspectos semântico-lexicais*, Master's thesis, São Paulo, USP.

Teixeira, José Aparecido (1938), *O falar mineiro*, São Paulo, Revista do Arquivo Municipal de São Paulo.
Teixeira, José Aparecido (1944), *Estudos de Dialetologia portuguesa: linguagem de Goiás*, São Paulo, Anchieta.
Toniolo, Ênnio José (1967), *A linguagem na depressão central gaúcha e a dos Campos Gerais*, Ponta Grossa, Minerva.
Toniolo, Ênnio José (1981), *Vocabulário de Tibagi*, Apucarana, Faculdade Estadual de Ciências Econômicas de Apucarana.
Toniolo, Ênnio José (2002), *Do presente ao passado: um olhar sobre o vocabulário de Tibagi*, Master's thesis, London, UEL.
Veríssimo, José (1888), *Scenas da vida amazonica*, Lisboa, Livraria Editora de Tavares Cardoso & Irmão.
Wouk, Miguel (1976), *Atlas lingüístico do Paraná. Primeira Etapa: Guaraqueçaba*, Estudos Brasileiros 1, 99–102.

Uli Reich and Ronald Beline Mendes

10 Sociolinguistics

Abstract: This chapter provides an overview of the current state of Brazilian sociolinguistic research. The complex nature of Brazilian society is itself seen as the main reason for the impressive wealth and diversity of work in this area throughout the country. We identify some of the most important centers, sketch some theoretical particularities of Brazilian sociolinguistics, and describe several large-scale projects. Some issues are traced from their initial dialectological descriptions to Labovian quantitative approaches and beyond to indexical fields in urban societies, which set the stage for future research.

Keywords: sociolinguistics, Brazilian Portuguese, language variation, language change, variable rules, parametric sociolinguistics, indexical fields, urban sociolinguistics, third wave sociolinguistics

1 Introduction

Any attempt to offer a synthesis of the social reality of Portuguese in Brazil and the associated linguistic research is a vast undertaking. Brazil is an enormous country, whose formation was configured by peoples from the four corners of the world, these from both forced and deliberate arrivals, and its demographic reality has constantly changed in waves of internal migration (see also ↗2 The social history of Brazilian Portuguese and ↗3 The history of linguistic contact underlying Brazilian Portuguese). Its different geographic regions are thus very diverse—economically, culturally and, consequently, linguistically. Its large cities, some of which are amongst the most highly populated megalopolises in the world, condense this diversity locally and project it onto scales of social values which are in constant renegotiation. Notwithstanding the many challenges that it poses, such complexity offers the perfect scenario for the study of language variation and its social meanings, as well as language change. In this chapter, we provide a concise overview of sociolinguistic research in Brazil, and although our description seeks to offer a representative view, it certainly cannot claim to be exhaustive. Indeed, we might go so far as to say that what we describe here constitutes merely the tips of several icebergs, this because even if we were to overlook the dialectological beginnings (see section 2.1) and thus only consider work that generally followed Labov (1966), it would be impossible to cover the full spectrum of sociolinguistic research that has developed in the country and which focusses not only on the production of Brazilian Portuguese (BP) variables, but also on perceptions and attitudes toward its varieties and other languages (see, for example, Bugel 2006), as well as studies in accommodation and language contact—both between varieties of

https://doi.org/10.1515/9783110405958-011

BP (Lima/Lucena 2013) and between BP and other languages (see ↗3 The history of linguistic contact underlying Brazilian Portuguese). We will focus mainly on research into production, that is, work dedicated to the description of variable uses in several varieties of BP and to explanations of social and linguistic embedding of variables; but we will also refer to some work on sociolinguistic perception, vis-à-vis the indexical fields of variants and interest in the social meaning of variation. In particular we will provide insights into work on grammatical variation, doing so by tracing some of the most salient variables from the dialectological beginnings of their description to their formulation as variable rules within the Labovian framework and beyond.

Although there are some aspects that can be said to unite BP despite the size and heterogeneity of the country (for example, see section 2.2 for our thoughts on the dichotomy *popular*/*culto* 'popular/cultivated'), it would be impossible to cover in a single chapter all specific instances of variationist analyses, ranging from the Portuguese spoken in each of the three southernmost states, to the varieties spoken in central Brazil and south-eastern cities, to those spoken in the North and the Northeast. Each of these regions has, so to say, their own "traditions", in terms of trends in variationist studies (for instance, the grammatical or phonetic variables most commonly analyzed). Therefore, our approach has been to cover people and traditions in one section (2.3.1), separate from the section in which we highlight a few of the most noteworthy variables studied in Brazilian sociolinguistics. Of course, many people, institutions and lines of research had to be left out, and this is not to be understood as implying any kind of bias, but simply as a reflection of the inescapable limitations of an overview of this kind. However, the focus on variationist sociolinguistics is indeed a delimitation imposed by the authors' own research histories, which do not include interactional sociolinguistics (Ribeiro/Garcez 1998) or discourse-oriented approaches (but see Pagotto 2004; Mendes 2018).

Finally, although this chapter concentrates on sociolinguistic studies that follow the Labovian paradigm (section 2.3), work that has departed to some extent from the macro-sociological perspective in the direction of more ethnographic ones (in the sense of Eckert 2012) is also noted (in section 2.5). Despite the necessarily condensed nature of the current chapter, we hope to offer a broad synopsis of this rich and fruitful area of linguistic research in Brazil, including a few considerations about its future direction.

2 Main issues in Brazilian Sociolinguistics

2.1 Taking the features to the city

As in many other regions of the world, some of the variables discussed in Brazilian sociolinguistics appear first in the literature as features of dialects. Work by Amadeu Amaral (1920) and Antenor Nascentes (1922) already identified some of the main do-

mains of variation in morphology and syntax in BP. Amaral (1920), for example, introduces the following features as characteristic of *Português Caipira* (roughly, the dialect spoken in the State of São Paulo outside its capital): (i) Lack of number agreement; (ii) Pronominal systems and subject-verb agreement; (iii) The oblique pronoun *mim* 'me' as a subject of embedded infinitives; (iv) The use of strong pronouns *ele(s)*, *ela(s)* 'he, she, they' in object position; (v) Relative clauses with a complementizer and a resumptive pronoun which does not undergo pied piping; (vi) Final positions of negative *não* 'no'; (vii) Extension of the preposition *em* 'in' to directions.

These forms coincide with the most prominent variables that would come to be studied in Brazilian sociolinguistics in correlation with social and linguistic factors. We comment on the sociolinguistic treatment of (i), (ii), (iv) and (v) in 2.3.3, where we also see that these are not particular features of the dialect of rural São Paulo, as Amaral presents them, but rather features of many dialects of Portuguese in Brazil and in other parts of the world where it is spoken, and indeed in spoken varieties of other languages.

The nature of the correlation of features of dialects with their social constraints in urban networks is far from trivial. In societies that have still not provided affordable systems of transport for the bulk of the population, and which lack an effective telecommunications infrastructure, like Brazil in Amaral's times, linguistic variation appears primarily as an indicator of distance in space (see ↗9 Dialectology and linguistic geography). Migration and the condensation of groups of people who speak many different dialects (and languages) in a big city diminish the importance of space and eventually lead to the exploitation of dialectal features for the construction of social meaning.

How do major theories of linguistic variation treat the relevant phenomena? In the Coserian projection of linguistic variation into varieties that are selected by place, social class, situation, and possibly other dimensions, the value of a feature in one dimension of variation entails its interpretation in another dimension in an ordered way (Coseriu 1980; Koch/Oesterreicher 1990; 2007). Thus, diatopically marked features would entail their interpretation as diastratically low, and would be bound to function in diaphasically informal registers. This order of feature interpretation, called 'chain of varieties' (German *Varietätenkette*), is not always as convincing as it appears in typical European societies, where standard varieties are built on the dialects of a single capital city where the elites of culture and power were located. In societies where standards are less clear, as in the case of BP, these correlations show themselves to be more complex. For example, *Carioca Portuguese* (the variety of Rio de Janeiro) and *Caipira Portuguese* (the variety of São Paulo, outside the capital) are more or less equally marked in a diatopical dimension, but while *Caipira* features are more commonly interpreted as "low at the diastratic level", *Carioca* features can receive higher values of social esteem. The attribution of these values always depends on the ideological position of the person attributing them. Hence, some people associate Rio de Janeiro with beauty, music and a liberal life-style, while others associate it with violence

and lack of efficiency, and differences can influence linguistic behavior (see Ploog/Reich 2005; 2006; Reich 2009 for attempts to integrate structural, discourse-pragmatic and socio-indexical restrictions theoretically).

Labovian Sociolinguistics does not propose a general rule for the ordered interpretation of linguistic features and constructions at different levels of variation. Rather, quantitative methods show the constraints of variation by meaningful social and linguistic variables, including covariation with stylistic and pragmatic categories in the first place, and leave space for further investigation into the particular socio-indexical fields they construct. We believe that this is a more phenomenological approach, and affords us a better means of getting to grips with the realities of these interactions than Coseriu's well-known chain of varieties.

2.2 The *Português Culto*/*Português Popular* dichotomy

Brazil's demographic history is one of forced migration that led to the breathlessly rapid urbanization of the country. Fleeing from droughts and poverty, northeasterners migrated in their millions to the southeast, settling in sparkling new cities and providing labor for the industries there. Taking the features and rules of their spoken dialects with them, they multiplied the plurilingual experience of the large cities to form pools of features that were exploited for the linguistic construction of social meaning. Bortoni-Ricardo's (1985; 2011) sociolinguistic studies of a satellite city of Brasilia constitute, arguably, the body of work which has most directly addressed this line of argumentation, although studies such as Ferrari (1994; 2008) for Rio de Janeiro, and Oushiro (2015) for São Paulo, also contribute to a body of research that has helped us to understand the meaning of linguistic urbanity, that is, the *popular*/*culto* distinction.

A crucial distinction, one that pervades the sociolinguistic literature on Portuguese in Brazil, maps words, features, rules and constructions in terms of two basic varieties of Portuguese: *Português Culto* and *Português Popular*. In many ways it appears to be similar to Ferguson's (1959) "high" and "low" varieties, with the difference that it is based more directly on educational than economic variables. Its impact, especially in historical linguistics, is impressive in Brazil (see ↗1 Brazilian Portuguese linguistics: an overview and ↗2 The social history of Brazilian Portuguese). Rodrigues (1987, 71–73) relates its rise to an operational concept used in large projects on the linguistic realities of Latin America directed by Lope Blanch (1967) in the Spanish speaking countries, and by Nelson Rossi (UFBA), Ataliba Teixeira de Castilho (UNICAMP) and Dino Preti (USP) in Brazil (2.4.1). According to Rodrigues (1987, 71), Lope Blanch identified the urban, educated speech of the capital cities as the most important variety to be studied, this eventually being seen as motivated purely by practical realities, in that it seemed impossible to tackle all varieties due to the lack of available linguists to carry out all the necessary empirical

work. Having defined the *culto*-varieties as the systems of Spanish spoken by people who had studied at a university, the lack of such an education served to define the alternative case, the *habla popular*. This model was followed by the linguists at the time by addressing the *Norma Urbana Culta* in their projects. Thus, the rise of the *culto*/*popular* distinction was actually based on practical decisions rather than on theoretical arguments. This dichotomy certainly reflects a major characteristic of Brazilian society, since a bipolar social structure, formed over centuries of colonialism and slavery, still divides the country today and places some Brazilians closer to advanced standards of education and culture than others. However, while it certainly serves well for a bird's eye view on linguistic variation of Portuguese in Brazil, and does indeed label with a high degree of approximation the major trends of change within BP, we believe that it cannot inform our understanding of the social meaning of linguistic variants in everyday speech in any substantial theoretical way.

Important aspects of linguistic variation are lost if we restrict our view of variation to the statistical distribution of certain linguistic facts within major social classes. For one thing, many of the features we observe in sociolinguistic interviews relate to universal pragmatic restrictions on linguistic forms in the different conditions of communication and in different types of discourse, rather than to social variables. Even more important for the exploration of the social meaning of linguistic variation is the notion of style, which is reduced to a binary opposition of "formal" vs. "informal" in many studies. Classical Quantitative Sociolinguistics seeks to express the probability of linguistic variants occurring in the language of a specific social class, but it does not elaborate on methods and concepts for the study of the construction of social meaning in everyday speech. The importance of style, alongside major social categories, has been recognized as an important aspect of linguistic variation in sociolinguistics since Labov (1966; 1972a; see also Bell 1984), but it has become the center of attention only recently. We find the most promising concepts for this field of research in Eckert (2000; 2008), who combines ethnomethodological reasoning with quantitative methods. In Brazil, studies in this field of research are at an early stage, but have brought about important insights into the way Brazilians use the variability of features from all levels of linguistic description to project their *persona* linguistically in different interactional settings (see below 2.5).

We might venture to go a step further in our reasoning about historical languages, systems and variation in social settings. If we map linguistic features according to the criterion of frequency of covariation with social macro-categories to subsystems of a historical language that are generally recognized as *popular* and *culto*, we effectively move an abstract step away from actual linguistic behavior, since most speakers use features of both varieties, depending on their social intentions. If we take the two varieties as substantial instances of linguistic competence, this would lead us to analyze utterances with features and rules attributed to both varieties as code-switching triggered by social and situational intentions and conditions. We believe that this analysis would imply an unjustified identification of the heuristic categories *Português Cul-*

to/Português Popular as instances of linguistic competence. Rather, we would be inclined to take information about the social indices of individual words, features, rules and entire constructions as part of the structural description of such elements, just as variable rules were conceived of in the theoretical foundation of Quantitative Sociolinguistics (Labov 1972a; Cedergren/Sankoff 1974). At the level of competence, then, we are not faced with two systems, but rather an additional, socio-indexical layer in the structural description of the elements of a polymorph, rhizomatic system that resembles historical Portuguese only loosely.

2.3 The extended Labovian paradigm

In Brazil, Labovian sociolinguistics has generated a very robust scientific community. We believe that the conjunction of two social and academic facts led to this. First, the late but rapid expansion and modernization of the Brazilian university system was characterized by an interest in innovative scientific frameworks. In the late 1970s and the 1980s, several Brazilian doctoral students went to the US to study either Generative Syntax or Quantitative Sociolinguistics. On returning to Brazil, they introduced these theories, with their accompanying methodologies, in Brazilian Linguistics departments.[1] Indeed, refreshingly new currents of research from outside the country were generally welcomed in a generation that was starting to find its way out of the lethargy of a society shut off from the wider world through Latin America's longest military dictatorship. The Labovian method of charting linguistic variation by social categories with a quantitative methodology was a useful tool for a scientific community eager to get to grips with its object of study, that is, the reality of Portuguese in the largest country on the continent. Spoken BP was perceived as hidden behind a curtain of prescriptive rules inherited from the grammatical tradition of a colonial power that anthropophagist modernism had already devoured in the first half of the 20th century.[2] In the 1970s and 1980s, Brazilian Linguistics still felt the need to discover the language that it sought to study. A second factor which contributed to the success of Labovian sociolinguistics can undoubtedly be seen in the configuration of Brazilian society itself. Broadly speaking, from the 1950s to the 1980s, massive waves of migration from the poor northeast to the large industrial cities in the southeast brought rur-

1 See ↗1 Brazilian Portuguese linguistics: an overview for more on the historiography of Brazilian Linguistics as a whole.

2 One of the earliest supporters of the project to study BP in the form that Brazilians actually speak it in everyday life, rather than in the form of examples drawn from literary sources, was Mário de Andrade. He had the ambitious project to write a *Gramatiquinha da Fala Brasileira* 'Little grammar book of Brazilian Speech', based only on spoken BP, and, very interestingly, without any "rules", taking stock only of "constancies"; see Pinto (1990), where the original text of the *Gramatiquinha* is published.

al dialects, which had developed their shape far removed from any prescriptive institution, to an increasingly urban society within rapidly expanding megalopolises.

2.3.1 Places, people and traditions

In Rio de Janeiro, Anthony Naro founded the school of sociolinguistics that led to the *Universidade Federal do Rio de Janeiro* (UFRJ), which became one of the most notable centers for sociolinguistic research in Latin America. Naro supervised the doctoral dissertations of Maria Luiza Braga, Maria Cecília de Magalhães Mollica, Marta Scherre and Nelize Pires de Omena, as well as the master's theses of Sebastião Votre, Dinah Callou, Charlotte Emmerich, Giselle Machline de Oliveira e Silva, Maria Carlota Amaral Paixão Rosa and Dante Lucchesi, among many others. All these scholars have played decisive roles in the development of Brazilian Sociolinguistics throughout the 1980s and 1990s, carrying out individual and collective research themselves, as well as supervising the projects of their own students. The departments of Linguistics and Portuguese continue the strong sociolinguistic tradition at UFRJ today, with different theoretical approaches. As examples, we will briefly note three of these. The historical interests of the research groups headed by Dinah Callou and Célia dos Santos Lopes encompass concepts of German Romance Philology. In a true Brazilian spirit that combines Generative Syntax with Labovian methodology (see 2.4), Maria Eugênia Lamoglia Duarte (1986; 1989; 1993; 2000) investigates the use of subjects and objects (see ↗6 Historical syntax for details). Meanwhile, Lilian Ferrari's work is based on social networks and draws on concepts of Cognitive Linguistics (Ferrari 1994; 2008). Thus, the *fundão*, as UFRJ is called in Brazil due to the artificial island on which it was built, continues to be a very important center for Brazilian sociolinguistics.

In São Paulo, the return of Fernando Tarallo from Philadelphia, where he studied with Gillian Sankoff, had a major impact on the linguistic roadmap of Brazilian linguistics. His work is deeply dedicated to linguistic change, both theoretically and from the perspective of the history of BP. In his short career thus far, his theoretical and methodological skills and his noted enthusiasm has inspired a great many research projects on syntactic variation and change. He proposed the idea that Generative Linguistics and Quantitative Sociolinguistics are research paradigms that can be integrated into a *Parametric Sociolinguistics*. We comment on this theoretical particularity of Brazilian sociolinguistics below (see 2.3.2).

In Brasilia, Stella Maris Bortoni-Ricardo founded a group on the lines of her own research into one of the most Brazilian of all sociolinguistic topics, the impact of migration to the cities on linguistic variation and its consequences for education. She takes as a case the unique construction of a brand-new capital, Brasilia, which brought together rural dialects that developed at some distance from cultural institutions. These dialects changed and merged to form new "rurban" dialects that structure the social networks in the capital. She published her doctoral dissertation first in Eng-

lish (Bortoni-Ricardo 1985), but we encourage the reader to consider the revised and updated Portuguese edition (2011) as well as the compilation of articles she edited with Vera Freitas and Ana Maria Vellasco (2010).

Many more centers of sociolinguist research emerged throughout the country and made significant contributions to the study of Portuguese across the huge multidimensional space of Brazil's social differences. The VARSUL project (*Variação linguística urbana da Região Sul*) is based within a network of several federal universities in the Brazilian South and provides a rich database for the study of social and local variation in this area (see Colischonn/Monaretto 2012). Dermeval da Hora organized the VALPB (*Variação Linguística no Estado da Paraíba—fases I–IV: variação*, *estilo*, *atitude e percepção*), which works principally on the study of phonological variables (Hora 2004; 2009). Dante Lucchesi's *Projeto Vertentes* has the aim of studying the reality of popular Portuguese in Bahia, thus feeding into the theory of an Afrobrazilian vernacular that would give rise to the *Portuguese Popular* (Lucchesi 2001). It is impossible to cite all relevant projects here, but see Biondo Salomão (2011) for a catalogue of 48 groups and projects working on variation within BP.

2.3.2 A particularly Brazilian blend in linguistic theory: Parametric Sociolinguistics

As we mentioned above, Brazilian sociolinguistics shows a unique capacity to integrate theoretical concepts from different strands of linguistic research, ranging from functionalist approaches and cognitive linguistics to generative models. Integrating the latter is quite unusual, given the antagonistic ideas about the object of study of "linguistics proper" voiced by its main proponents and many of their followers. While Labov argues that linguistics needs to describe and explain the rules and constraints of speech in everyday life (Labov 1972a, 187), the position of Chomsky (1965; 1981; 1995) is that we must abstract away from speech if we want to understand what language is in its essence. The corresponding methodologies, that is, empirical data collection and statistical analysis versus introspective judgments of grammatical well-formedness and the projection of particular constructions onto universally available abstract forms, seem to make any attempt to combine these wildly differing schools futile. In Brazil, some scholars do indeed follow the tenets of a specific school of linguistics exclusively, but many others try to find ways of building theoretical bridges and integrating the particular advantages of both theoretical architectures. Tarallo/Kato (2007, [1]1989) is the theoretical article that established the particular practice of addressing variation that has led to many studies on Brazilian Portuguese being substantially different from other areas of sociolinguistic research. The authors take the controversies that exist between generativists and variationists as instances of the old opposition between empiricists and rationalists, which in turn they trace back to the antagonism between neo-grammarians and empirical scholars in the 19th century. As a result, they argue that these different views are complementary rather than antago-

nistic. They point to the (then) recent change in Generative theory from the rule-based Standard Theory (Chomsky 1965) to the Theory of Principles and Parameters (1981) model, and demonstrate that a parametric approach to syntactic variation is easily combinable with the results of a probabilistic view of variation in recorded data of real speech. In their view, the features of parametric change across different syntactic domains will be observable in linguistic performance in different diachronic steps, resulting in the synchronic impression that different features appear in constellations different from those predicted by the theory. The probabilistic approach to diachronic data would lead to the incremental completion of the whole parameter being visible over time. Briefly, while the probabilistic approach can translate abstract theoretical predictions into linguistic reality, the rational notions of structural relations across syntactic domains can guide the variationist's view of linguistic conditions for variation. In the authors' own words: 'Both theories are big. Each one grows inasmuch it feeds on the other'[3] (Tarallo/Kato 2007, [1]1989, 38). This theoretical position usefully and productively pervades a good deal of sociolinguistic research in Brazil.

2.3.3 Some prominent variables—Syntax and morphology

Just as elsewhere, Quantitative Sociolinguistics in Brazil began with studies which looked mainly at phonological variation. In many cases, morphological and/or syntactic forms turned out to be a significant predictor, and in some cases the relation is seen to become upturned during the course of research: that which was the predictor becomes the variable under study, and phonological variables come to be seen as the set of linguistic conditions. Syntactic variation soon entered the spotlight of sociolinguistic research in Brazil. Here, we focus on three prominent morphological and syntactic variables, since they show particular characteristics of Brazilian Sociolinguistics better than does research on lexical or phonological variation. To our knowledge, sociolinguistic research in prosodic phonology (rhythm, metrical phonology and intonation) has not yet begun, despite the readily observable differences in intonation, metrical patterns and rhythmic forms across different communities. We have chosen to present selected elements from a small number of studies, rather than to cite many studies without any accompanying description.

Fernando Tarallo's impact on Brazilian Sociolinguistics has earned him a reputation that makes it almost impossible to discuss this topic and not begin with his dissertation on variation of relative clauses (Tarallo 1983). In Brazil, relative clauses can take three forms, the standard variant with a wh-constituent and pied piping (1a), a variant with a strong resumptive pronoun (1b), and a variant without such a pronoun (1c):

3 "Ambas as teorias são grandes [...] Cada uma cresce à medida que da outra se alimenta."

(1) a. o cara com quem eu falei
ART guy with who 1SG talk-PST.1SG
b. o cara que eu falei com ele
ART guy REL 1SG talk-PST.1SG with 3SG
c. o cara que eu falei
ART guy REL 1SG talk-PST.1SG
'The guy I talked to'.

Looking quantitatively at interviews, data from the media, and diachronic corpora, he finds evidence for the claim that lower class speakers prefer relatives of the (1c) type and that "spontaneous style" favors the (1b) type. Regarding these differences, he postulates the competition of two grammatical systems in BP. Diachronic analysis leads him to the postulation of a harsh syntactic change at the beginning of the 19th century that would involve also the syntax of subjects and objects.[4] Overall, this thesis set the stage for research which would focus mainly on parametric syntactic change, although its main proposition of a parametric correlation between the loss of the pro-drop features of the subject and allowing empty categories for arguments in object position does not seem to have survived to the present time.

Let us look at variation in two other domains of agreement and cite the work of two scholars that are representative of Brazilian Sociolinguistics in the late 1980s and early 1990s. Martha Scherre's work on nominal plural agreement (Scherre 1978; 1981; 1996) shows exactly the inversion of perspective on the facts of co-occurrence just mentioned. To begin with, this variable is conceived of as phonological in nature, but with morphosyntactic covariation, a perspective that is inverted over the course of research.

The variable shows three variants. (2a) is the canonical form with plural *-s* suffixed to all elements of the nominal group, (2b) shows it on the two first, and (2c) only on the first element:

(2) a. os meninos malucos
ART.M-PL boy-M-PL crazy-M-PL
'The crazy boys'
b. os meninos maluco
ART.M-PL boy-M-PL crazy-M
c. os menino maluco
ART.M-PL boy-M crazy-M

4 The diachronic facts, however, are controversial, since the analysis did not take into account the possibility of a change in discourse traditions, allowing for forms of spoken BP to appear in written texts of communicative distance; see ↗6 Historical syntax for more arguments in this sense.

Scherre's work illustrates the importance of a concept introduced by Lemle/Naro (1977) and Naro (1981) that proved to be decisive for other studies. We can take it as an early glimpse *avant la lettre* of a sociocognitive approach to linguistic variation (see Kristiansen/Dirven 2008; Krefeld/Pustka 2010; Caravedo 2014). The *principle of phonic salience* states that phonological oppositions show different degrees with respect to their perception and that these differences contribute to determining whether a grammatical rule applies or not. Scherre demonstrates that the greater the phonetic difference between the singular and the plural forms, the more likely it is that a speaker will use the canonical form, as in (3a). Differences can range from simple suffixation (3a), suffixation with metaphonic processes (3b), or even both, combined with stem modification (3c):

(3)	a.	a	casa	as	casa-s
		ART.F	house	Art.F-PL	house-PL
		'the house'		'the houses'	
	b.	o	[o]vo	o-s	[ɔ]vo-s
		ART.M	egg.M	ART.M-PL	egg.M-PL
		'the egg'		'the eggs'	
	c	o	pãozinho	o-s	pãezinhos
		ART.M	bread-DIM	ART.M-PL	bread-DIM-PL
		'the bread roll'		'the bread	rolls'

Another important linguistic variable here is the topological position of a given element in the nominal phrase with respect to the head of construction that is identified as the noun. The greater the distance between head noun and dependent elements, the more likely it is that the dependent element will not show plural morphology. Positions to the left of the head noun favor plural morphology, and head nouns are more frequently marked for plural if they are placed in initial positions.[5]

This set of linguistic conditions is correlated with extralinguistic macro-categories, namely the classic three: sex, education and age. While age seems to play no significant role, education and sex are seen to yield significant probabilistic effects. Speakers with 1 to 4 years of schooling show a relative weight of 0.35 in applying plural morphology, speakers with 5 to 8 years 0.53, and speakers with 9 to 11 years 0.66. The probabilistic factor for the application of the plural for male speakers is 0.41, and for female speakers it is 0.58 (Scherre/Naro 1998, 519).

Agreement is also involved in the second morphosyntactic variable to be discussed here. It is related to an important change in the pronominal system that came

5 An obvious alternative for this NP-based approach would be to take the determiner as the head of construction in a DP analysis. We are not aware of any studies taking that approach. Especially in light of the facts on the relevance of topological position, we believe that this might be a promising avenue for future work.

about via the grammaticalization of the noun phrases *vossa mercê* and *a gente* to the personal pronoun *você* for second person singular (Vitral 1996), and *a gente* for first person plural (Zilles 2005; Lopes 2003; 2007). Both compete with canonical forms and show different forms of agreement (or non-agreement) with verbal morphology if they surface as subjects, as we can see in (4a–d) for first person plural, and in (4e–h) for second person singular:

(4)	a.	nós	vamos	b.	nós	vai
		we	go-3PL		we	go-3SG
	c.	a gente	vamos	d.	a gente	vai
		we	go-3PL		we	go-3SG
	e.	você	vais	f.	você	vai
		you	go-2SG		you	go-3SG
	g.	tu	vais	h.	tu	vai
		you	go-2SG		you	go-3SG

It is important to note that all forms except (4e) are in use in Brazil. Sociolinguistic research focusses on different aspects of this constellation, (i) the selection of the pronoun (Lopes 2003; ↗6 Historical syntax), (ii) the possibility of dropping a pronoun (Duarte 1993; 1995; Duarte/Cavalcante 2008), and (iii) the pronoun and the morphological form of the verb (Rodrigues 1987; 2000; 2004), as in (5):

(5)	nóis	fazia	lavora	é.
	we	make-PST.3SG	farm labor	indeed
	'We did farm labor, indeed'			
			(Rodrigues 2004, 134)	

This phenomenon has been studied in many projects, dissertations and articles (Lemle/Naro 1977; Naro 1981; Guy 1981; Bortoni-Ricardo 1981; 1985; Baxter/Lucchesi 1993; Mello 1996, among others). We take Rodrigues (1987; 2000; 2007) here as an example that shows the way orthodox Labovian Sociolinguistics worked well in Brazil in the 1980s. She studies subject-verb agreement in first person plural and third person plural in a corpus that she recorded in the *favelas* of the *Jardim do Carombé* neighborhood in the northern outskirts of São Paulo. She classifies her 40 informants following the social categories of sex, level of education, age and origin, contrasting speakers born in the city of São Paulo with migrants from three different rural areas. The linguistic variants for first person plural (1PL) are the phonological shape of the canonical form, distinguishing verbs with antepenultimate stress from verbs with penultimate stress (*falávamos*/*falamos*), and the morphosyntactic form of the subject that could be a lexically expressed overt pronoun, an empty category, or not a pronoun. For third person plural 3PL, the linguistic factors considered are the topological posi-

tion of the pronoun (directly before the verb, directly after the verb, or separated from the verb), and again its phonological form, building on the concept of "phonic salience" from Lemle/Naro (1977). The verbs were classified on a six-point scale with values of increasing phonic salience:

(6) phonic salience
(6.1) only nasalization or change of quality of a final, metrically weak vowel: *fala*/*falam* ['fala/'falãw̃],
(6.2) difference in final metrically weak vowel that is added to the stem: *faz*/*fazem* [faz/'fazẽỹ],
(6.3) quality difference in metrically strong vowel: *dá*/*dão* [da/dãw̃],
(6.4) addition of a full syllable and change of quality of a vowel in a metrically strong position: *falou*/*falaram* [fa'lo/fa'larãw̃],
(6.5) change of vowel in the stem and addition of a full syllable: *fez*/*fizeram* [fez/fi'zerãw̃],
(6.6) suppletive forms: *é*/*são* [ɛ/sãw̃];

This set of morphophonological possibilities has the highest predictive power as a condition for both variables: The higher the phonic salience, the more probable it is that speakers will apply a rule for agreement, both for 1PL and for 3PL. All social factors proved to be significant. Both education and origin are good predictors. The analysis of the data confirms the overall expectation that education and urbanity (born in the city of São Paulo) favor the use of agreement morphology on verbs. It is the variable "sex" that leads the author to a theoretical innovation in Brazilian sociolinguistics. Discussing the fact that men favored the standard forms relative to women, particularly for 1PL, Rodrigues (1987; 2000) develops a concept that points towards the role of perception in social variation and social networks: *Social salience*. In the communities under study, women's networks are more associated with community-internal relations, while men show more mobility and thus are associated with a higher degree of contact outside the *favelas*. In such contexts they would perceive certain forms as more stigmatized than others: thus, *nós* with verbal morphology, which the standard variety recommends for 3SG, is socially more salient because its forms are used more often in public discourse, giving them an emblematic function as an indicator of the social construction "illiterate from the northeast". We discuss further socio-indexical values of this variable below (2.5.1).

Duarte (1986; 1989) works on a very salient syntactic variable in the context of other Romance languages. In object position, Brazilians employ either full NPs, the canonical clitic pronouns *o(s)* and *a (s)*, the strong pronouns *ele(s)* and *ela(s)*, or no phonetic realization of the corresponding argument, interpreted here as an empty category in the sense of Chomsky (1981). The corpus she uses consists of sociolinguistic interviews with 50 speakers from the city of São Paulo and recordings from the TV. Speakers are categorized following the independent social variables used as predic-

tors, the three levels of Brazilian education and three age groups. The linguistic variables are morphological, syntactic and semantic in nature, the most relevant being the semantic feature [+/- animated] that strongly favors the use of strong pronouns, and the syntactic condition of the form of the antecedent, ruling out the use of the strong pronoun with sentential antecedents. The social predictors show clear restrictions for the clitic pronoun, this being absent in the speech of the youngest group (15–17 years), and on the use of the strong pronoun, which occurs significantly less in the speech or more educated *Paulistanos* (those with a university degree).[6]

Duarte (1993; 1995) and Cavalcante/Duarte (2008), among many others, follow Tarallo (1983) and Tarallo/Kato (2007, [1]1989) in applying the theoretical architecture and empirical methodology of Quantitative Sociolinguistics in the sense of Weinreich/Labov/Herzog (1968) to study change in the Brazilian way as described in (2.3.2), above. They integrate the formal apparatus of Generative syntax in establishing sets of linguistic conditions for variation and in testing for predictions of Principles and Parameters Theory (Chomsky 1981; 1995). The domain under study is the syntax of subjects. Duarte (1993) introduces two features of the pro-drop parameter, rich or poor verbal morphology, and the possibility of subject inversion, as predictors of another feature of this parameter: the possibility of dropping the subject pronoun. She applies this perspective in a quantitative study of a corpus containing theater plays from the early 19th to the late 20th centuries, and finds that the presence of subject pronouns increases over time, while verbal morphology is reduced to a paradigm with only two oppositions (*eu falo*, *você fala*, *ele fala*, *a gente fala*, *vocês fala(m)*, *eles fala (m)*) and post-verbal subjects decrease (see ↗6 Historical syntax for full documentation). This, of course, confirms the validity of the *Parametric Sociolinguistics* approach very felicitously.

Other theoretical enrichments of Brazilian Sociolinguistics come from functionalist and cognitive approaches. The historical development of *a gente* from a noun phrase to a pronoun for 1PL has been addressed in diachronic research by Lopes (2003). At that time, it was already possible to conduct studies in real time, with the same communities of the first sociolinguistic recordings in the 1970s (see 2.4.1). Thus, Lopes was able to build on three methodological axes of data collection: A study over a long time-frame tracing the forms back to the 13th century, apparent time studies with different age groups, and real time studies. Her theoretical foundations combine

6 However, Reich (2002) shows in a corpus study of 20 *Paulistanos* between 25 and 35 years old, all with university degrees, that these social and semantic restrictions do not hold across all communities. In this corpus, unanimated referents are readily expressed with strong pronouns, the only restriction being syntactic: strong pronouns in object position do not allow for predicates as antecedents: *Tentei [parar de fumar]$_i$, mas não consegui (*ele$_i$)*. We believe that this difference is due to the relaxed situation of the recordings, which were carried out with close friends, in contrast to the rather anonymous situation of a typical sociolinguistic interview. Other possible explanations are, of course, change or dialectal differences.

Labovian Sociolinguistics with grammaticalization theory. She observes the categories number, gender and person in occurrences of *a gente* in her corpora, and shows the loss of number specification, the loss of gender, and the retention of the morphological category as 3SG, interpreting her findings as an ongoing process of grammaticalization. Lopes/Callou (2004) compare two corpora recorded in the 1970s in terms of the same variable, the *Norma Urbana Culta* Corpus for Rio de Janeiro and the CENSO-Corpus, documenting the speech of less-educated speakers.

2.4 Large projects

2.4.1 NURC/GPF

The project to study the urban standard varieties of five major cities (São Paulo, Rio de Janeiro, Salvador da Bahia, Recife and Porto Alegre), as mentioned earlier, was established in order to complement a comparable project in the Spanish-speaking Americas (see also ↗1 Brazilian Portuguese linguistics: an overview). Its main promotors in Brazil were Nelson Rossi (UFBA), who initiated the project in the country, and later Ataliba Teixeira de Castilho (UNICAMP), who extended the project to a nationwide study in spoken BP (see below and ↗1 Brazilian Portuguese linguistics: an overview). The regional coordinators were José Brasileiro Villanova for Recife, Nelson Rossi and later Jacyra Andrade Mota e Suzana Marcellino Cardozo for Salvador da Bahia, Celso Cunha Isaac Nicolau Salum and later Dinah Isensee Callou for Rio de Janeiro, Ataliba Teixeira de Castilho, Isaac Nicolau Salum and later Dino Preti (USP) for São Paulo, and Albino de Bem Veiga for Porto Alegre. The project documented speakers with university degrees in three discourse types. The *diálogos entre informante e documentador* (DID) consist of classic sociolinguistic interviews with fixed topics, the *diálogos entre dois informantes* (D2) involved conversations on fixed topics without the monitoring participation of a linguist, and *elocuções formais* (EF) were recordings of lectures and talks at universities. The project was very ambitious, recording a total of 1,570 hours and 40 minutes of speech by 2,356 informants in 1,870 recording sessions. Parts of the recordings were transcribed and published (see Castilho 1990 for the history of the project, with an extensive bibliography). The first analysis appeared in Preti/Urbano (1990), but the systematic quantitative description of the linguistic facts at all levels of grammar took place within the follow-up project on spoken BP. The project on the *Gramática do Português Falado* has arguably been the most ambitious project in the history of Brazilian Linguistics, bringing together linguists from all theoretical orientations to analyze and discuss the data from NURC. Its main promoter was Ataliba Teixeira de Castilho, who had the linguistic vision and personal charisma to make the project possible. Many of the contributions are written in the spirit of *Parametric Sociolinguistics*, showing fine-grained analytical concepts derived from formal frameworks, these coming mainly from Generative Linguistics, and incorporating theoretical insights into

variation and the methodological skills needed to work with empirical data. As noted above, the close interaction of linguistic theories which elsewhere were held to be antagonistic is perhaps the most original contribution of Brazilian Linguistics to linguistic theory-building.

The project led to eight volumes of original contributions, four volumes with detailed articles on different aspects of the grammar of Portuguese spoken in Brazil, and finally a reference grammar authored by Castilho (2010).[7] Many of the contributors continued to work on the *Projeto para a História do Português Brasileiro*, using the theoretical and methodological foundations that had been developed therein.

2.4.2 Censo/PEUL

This was perhaps the most sociolinguistically oriented of all the large-scale projects in Brazil, and it is a story of genuine success in linguistic research. Moreover, it can be considered as the first major step in the description of urban BP, as well as being perhaps "the most social" of the sociolinguistic projects in Brazil, since it was conceived of as complementing the NURC project, seeking a more complete picture of the reality of Portuguese in Brazil by targeting the grammar of less educated Brazilians. It was initiated by doctoral students of Anthony Naro's, most of these from UFRJ but some also from other universities: Marta Scherre, Sebastião Votre, Giselle Machline de Oliveira e Silva and Charlotte Emmerich. This group was completed by Nelize Pires de Omena, Maria Cecilia Mollica, and Jürgen Heye, and subsequently by Maria Luiza Braga and Maria da Conceição Paiva. Gregory Guy and eventually Gillian Sankoff also worked directly with them (Paiva/Scherre 1999). By 1980 they had begun the *Censo da Variação Linguística no Estado do Rio de Janeiro*, which recorded an initial database consisting of 48 hours of speech from 48 adult informants categorized along the familiar parameters of sex, age and education, complemented later on with more recordings. In 1987, the *Programa de Estudos sobre o Uso da Língua* (PEUL) was founded and started its work on *Português Popular*. Unfortunately, it never found its way across the boundaries of Rio de Janeiro, and did not come to constitute a nationwide network like the NURC project. It would be of enormous interest to compare the sociolinguistic realities of different cities in Brazil, since such a comparison might reveal significant insights into the importance of migration and networks on variation. To date, PEUL has been the most robust collective project on the linguistic competence of less educated Brazilians. Theoretically, the group from Rio shows more affinities to functionalist and cognitivist linguistic theories than does the brand of sociolinguistics in São Paulo. The results of the project are published in Tarallo (1989), Silva/Scherre (1996), special

7 All publications are extensively referenced in ↗1 Brazilian Portuguese linguistics: an overview, and we have chosen to not repeat them here.

editions of DELTA (1989/2, 1991/1 and 2, 1992/2 and 1993/1) and *Language Variation and Change* (1991/1, 1991/3 and 1992/1 and 2), among many individual books, dissertations and articles.

2.5 New horizons

In addition to the definition of a variable ("alternative ways of saying the same thing", Labov 1972a), the concept of prestige is crucial to Labovian sociolinguistics, to the extent that, depending on the attention that one pays to speech, their position in social hierarchy is an important predictor of which variant is likely to be used. As we have seen, sociolinguistics in Brazil has widely adopted such a framework. This means that, along with a description of regional differences in terms of phonetic and grammatical variation patterns (from Southern and Southeastern, to Northern and Northeastern states and their cities), the prestige model has been utilized and tested again and again in Brazilian sociolinguistics, as references in the previous sections attests. Noun phrase and subject-verb number agreement are among the examples of extensively studied variables throughout Brazil (see 2.3.3 and Mendes/Oushiro 2012 for a further review of these studies).

Traditionally, Brazilian sociolinguistics has made wide use of the sociolinguistic interview as a tool, in the construction of samples that are random (e.g. Projeto ALIP, Gonçalves/Tenani 2008) or semi-random (e.g. Projeto SP 2010, Mendes/Oushiro 2012). However, there have been attempts at different methodologies in approaching a speech community. Drawing on Rodrigues (1987), Coelho (2006), for example, had to reconsider his aims in building a random, balanced sample of speakers living in Brasilândia, a slum neighborhood in the city of São Paulo. Interested in first-person plural verb-subject agreement (considering as variants *nós V-mos* 'we V-1p.pl', *nós V-Ø* 'we V-3p.sg', *a gente V-Ø* 'we V-3p.sg' and *a gente V-mos* 'we V-1p.pl', as in examples (3) above), Coelho (2006) realized, in his ethnographic observations, that among older speakers in his sample he was only able to include seamstresses that worked in a local cooperative, female day-care teachers, and members (mostly male) of the neighborhood association. On the other hand, among the younger, he could only sample speakers from the youth rehabilitation project (mostly women), the children drawn from the associations, or boys that spent most of their time playing soccer in the neighborhood field (the so-called *manos* 'bros'). Therefore, only two of these local subgroups include both male and female speakers, in such a way that for most categories, Coelho's (2006) sample could only include either male or female speakers, either younger or older. In his analyses, Coelho (2006) found that the pronoun *nós* 'we' is preferred by the *manos*, with nonstandard agreement (*nós V-Ø*). The pronoun *a gente* is relatively preferred by the day-care teachers. These results allow the interpretation that *nós V-Ø* works locally as a symbol of the *manos*' rebellion against the values of adults. Coelho (2006, 151) sees it thus as a case of "vernacular exaggeration"

which tends to be rejected by the other members of the community, especially those that do not want to appear affiliated to the "bad boy" image, which is how the *manos* are characterized in the neighborhood.

Another example of alternative methods of sampling is given by Freitag et al. (2016) in Sergipe (a state in northeastern Brazil). In order to address, among other variables, the affricate pronunciation of /t/ and /d/ that follow a palatal glide (as in [mujtʃʊ] *muito* 'much' and [dojdʒʊ] *doido* 'crazy'), Freitag et al. build a sample of conversations between speakers based on the proximity of their relationships, in a gradient that goes from "close" (when the interlocutors know each other and are part of a same circle of friends, and who talk to one another frequently) to "more distant" (when the interlocutors do not know each other and are talking to each other for the first time during the recording session). Each speaker interacted with four others (a man and a woman that they knew or were friends with, and a man and a woman that they didn't know). There was a total of 32 conversations of about 40 minutes each. Each pair of speakers could talk about different topics that they previously selected from a series of situations described on cards (which included bullying, public transportation, environmental issues, among others). So, the topics were not defined by or organized according to a script, as is usually the case with sociolinguistic interviews, and were not conducted by an interviewer. Freitag et al. (2016) show that the affricate pronunciation of /t/ and /d/ is significantly more frequent when the relationship between the interlocutors is close. Conversely, the affricate variant is avoided (and the non-affricate—as in [mujtʊ] *muito* 'much' and [dojdʊ] *doido* 'crazy'—is preferred) when the interlocutors don't know each other. Taking into account that the affricate variant is avoided in Aracaju (Sergipe's capital), while it is favored in its surrounding metropolitan areas, this is not simply a case of orientation towards central or peripheral areas of a city, but rather a case in which the nature of the relationship between speakers has an undeniable role —a hypothesis that could only be tested thanks to the type of sample collected.

Ethnographic approaches such as these, which try to refrain from macro-categories (typically social class, sex and age) and pay closer attention to locally defined micro-categories, constitute what Eckert (2012) calls "the second wave of sociolinguistics". According to her, despite their methodological differences, first- and second-wave sociolinguistics share an interest in language change, and even though second-wave research moves toward a vision of the speaker as a social agent, rather than as a passive reflex of their macro-categorization, it is in the third wave that we return to the importance of the social meaning of linguistic variation and focus on speakers' agency and speech styles as sources of social differentiation. Crucial to the so-called third-wave sociolinguistics is the notion that the social meanings of variants are themselves variables (Eckert 2008). Moreover, the indexical link between a linguistic form and a certain social meaning can be indirect (Ochs 1992; Silverstein 2003), in a way that, for example, *nós V-Ø* may function as an index of social class, whereas in Brasilândia, Coelho's ethnographic observation and analyses show that it is more directly linked to the local identity of "rebellious younger males".

Eckert (2012, 88) notes that third-wave sociolinguistic research is in its infancy. Indeed, in Brazil it is still very much in its infancy. To date, there are very few studies on the analysis of clusters of linguistic features employed by speakers in the construction of personae, which in turn are mapped onto styles (see, for example, Podesva 2006). One of the exceptions is Mendes' (2018) analysis on how 4 male speakers combine variable NP agreement and the pronunciation of nasal /e/ (as a diphthong or a monophthong, in words like *apartamento* 'apartment'), in order to approximate to or distance themselves from "macho" stances. However, in addition to work on attitudes towards Brazilian varieties and variants (e.g., Fraga 2009; Botassini 2011; Silva/Aguilera 2014), there has been some work on sociolinguistic perceptions, with the aim of discovering the indexical field of variants. One example here is Oushiro (2015), who used the matched-guise technique (Lambert et al. 1960; Campbell-Kibler 2006) to elicit listeners' reactions to audio stimuli that contained either tap or retroflex /-r/. These are the two most productive variants of coda /-r/ in São Paulo speech and, while the former is generally evaluated (by natives of the city) as "normal" and as a *Paulistano* feature, the latter is referred to as an accent feature, typical of those that come from the state's country side (*Caipira*, as we've seen above). However, even though the matched-guise experiment carried out by Oushiro (2015) confirms the association between these variants and those social meanings, it also shows that the participants' perceptions differ depending on whether they live in more central or peripheral areas of the city and on their origins. Those who live in peripheral areas and those who migrated to São Paulo from other cities in the state consider retroflex /-r/ to be just as *Paulistano* as the tap. Conversely, those who live in central São Paulo and migrants from other states tend to perceive the retroflex as a marked, non-*Paulistano* form.

Although noun-phrase number agreement is a widely studied variable in BP, it has never been looked at in terms of its indexicality (beyond the correlations verified in production studies). Inspired by production analyses that suggest that effeminate-sounding men favor the use of standard NP plural marking while masculine-sounding men favor the non-standard, Mendes (2016) discusses the results of a perception experiment which concludes that effeminate-sounding men are perceived as less effeminate when they use nonstandard NP plural marking, and that masculine-sounding men are perceived as less masculine (more effeminate) when they use the standard form of plural marking. The male speakers whose voices were used as stimuli in this experiment are also perceived, in the standard-guise, as more educated, more intelligent, more formal and pertaining to a higher social class (though not as more friendly). By exploring the notion of indirect indexicality, Mendes (2016) shows that while male listeners associate perceived level of education with perceived masculinity/effeminacy less indirectly, female listeners associate perceived level of education first with notions of intelligence and class, and then with masculinity/effeminacy. Such research (Oushiro 2015; Mendes 2016; Mendes 2018) indicates that Brazilian sociolinguistics is also experiencing "a return" to the social meaning of variation as a central focus for research (in the words of Eckert 2012).

3 Challenges for the future

According to Eckert (2012), the genesis of what she calls "third-wave sociolinguistics" was Labov's work in Martha's Vineyard, in which Labov (1972a, 38) concludes that the social meaning of centralized /aw/ and /aj/ is "positive orientation" towards the island. Many studies in sociolinguistics around the globe have been inspired by this seminal study, including later re-studies of the same community (e.g., Blake/Josey 2003). In Brazil, things were no different. Bieler da Silva (2015), for example, followed Labov's steps in her study of coda /-r/ pronunciation in Itanhandu, a small town in Minas Gerais. Located in the south of that state, an area where the retroflex pronunciation of coda /r/ predominates, Itanhandu stands out because fricative and tap /-r/ occur in local speech. This is attributed to the fact that people from Itanhandu often move to Rio de Janeiro and São Paulo, metropolises where these two pronunciations are the most common (and where the retroflex is infrequent). Bieler da Silva (2015) found that a negative orientation to Itanhandu (and a positive orientation to either Rio or São Paulo) had no correlation with the occurrence of tap and fricative /-r/ in the Southern Minas town. Instead, time spent in those cities is the variable that explains this variation: the longer the time away from Itanhandu, the more likely was coda /-r/ to be pronounced as a tap or fricative. These results differ from those found by Carmichael (2014) in her post-Katrina study of three locally salient variables of English in the New Orleans region: none of the variables correlates to whether an individual moved away or returned following the storm; here, it was orientation towards the place that predicted the variation. In Carmichael's words, her results "demonstrate that physical movement across spaces affects speech patterns less than one's relationship with a given place" (2014, 286). It is certainly not difficult to accept the proposition that how we conceive of ourselves in the world matters in terms of how we speak. However, we need more research to deal with such issues: What matters more, place identity (Johnstone 2004) or exposition/access to different variants (and their social meanings)?

This question is of the greatest interest in the study of variation within large cities, where mobility and migration are not exceptional. Take, for example, Oushiro's (2015) findings on the perception and production of /-r/ in São Paulo. In general, *Paulistanos* tend to evaluate retroflex coda /-r/ as a mark of non-*Paulistano* status and to believe that their speech presents no variation—that is, that their pronunciation is categorically a tap. However, in a representative sample of *Paulistano* speech (see Oushiro 2015, 35–40), 9,226 tokens of pronounced coda /-r/ were extracted and among these 28.3 % were represented by the retroflex variant; this is not in itself a small percentage, considering the generally negative social evaluation of the retroflex. One of the most interesting findings by Oushiro is that, although speakers' age does not predict this case of variation, there is an interaction between the area of residence (central vs. peripheral) and age: *Paulistanos* that live in the peripheral areas (and who tend to be from lower social classes) exhibit a pattern of change in progress toward the retroflex;

on the other hand, *Paulistanos* from central areas of the city (who tend to pertain to middle, upper-middle or higher social classes) show stability, with both older and younger people avoiding retroflex /-r/. This is just one example of how complex the patterns of variation in a large city can be. Perhaps with the advance of third-wave studies in Brazilian sociolinguistics, we might see the emergence of "big-picture" research questions with an interest in urban styles and identities that are defined more locally. In other words, considering the tendency of Brazil's scholars to combine theoretical perspectives, as we have seen above, the future of the field of sociolinguistics may witness a convergence of macro-sociological, ethnographic and stylistic approaches.

4 References

Amaral, Amadeu (1920), *O dialeto caipira*, São Paulo, O Livro.

Baxter, Alan/Lucchesi, Dante (1993), *Processos de descrioulização no sistema verbal de um dialeto rural brasileiro*, Papiá 2, 59–71.

Bell, Allan (1984), *Language style as audience design,* Language in society 13/2, 145–204.

Bieler da Silva, Mariane (2015), *Entre duas metrópoles: (-r) em Itanhandu*, Master's thesis, São Paulo, Universidade de São Paulo.

Blake, Renée/Josey, Meredith (2003), *The /ay/ diphthong in a Martha's Vineyard community. What can we say 40 years after Labov?* Language in Society 32/4, 451–485.

Bortoni-Ricardo, Stella Maris (1981), *A concordância verbal em português: um estudo de sua significação social*, in: Hildo Honório do Couto (ed.), *Ensaios de linguística aplicada ao português*, Brasília, Thesaurus, 79–101.

Bortoni-Ricardo, Stella Maris (1985), *The urbanization of rural dialect speakers. A sociolinguistic study in Brazil*, Cambridge, Cambridge University Press.

Bortoni-Ricardo, Stella Maris (2011), *Do campo para a cidade—estudo sociolinguístico de migração e redes sociais,* São Paulo, Parábola.

Bortoni-Ricardo, Stella Maris/Vellasco, Ana Maria de Moraes Sarmento/Freitas, Vera Aparecida de Lucas (edd.) (2010), *O falar candango: análise sociolingüística dos processos de difusão e focalização dialetais*, Brasília, UnB.

Botassini, Jacqueline Ortelan Maia (2011), *Crenças e atitudes linguísticas: Um estudo da relação do português com línguas de contato em Foz do Iguaçu*, Línguas e Letras 12/22, 65–84.

Bugel, Talia (2006), *A Macro- and Micro-Sociolinguistic Study of Language Attitudes and Language Contact. Mercosur and the Teaching of Spanish in Brazil, Doctoral thesis*, Urbana and Champaign (IL), University of Illinois Urbana-Champaign.

Campbell-Kibler, Kathryn (2006), *Listener perceptions of sociolinguistic variables: the case of (ing)*, Doctoral thesis, Stanford (CA), Stanford University.

Caravedo, Rocío (2014), *Percepción y variación lingüística: enfoque sociocognitivo*, Madrid, Iberoamericana.

Carmichael, Katie (2014), *"I never thought I had an accent until the hurricane": Sociolinguistic Variation in Post-Katrina Greater New Orleans*, Doctoral thesis, Columbus (OH), Ohio State University.

Castilho, Ataliba Teixeira de (1990), *O português culto falado no Brasil: história do projeto NURC/ Brasil*, in: Dino Preti/Hudinilson Urbano (edd.), *A linguagem falada culta na cidade de São Paulo,* São Paulo, TAQ/Fapesp, 141–202.

Castilho, Ataliba Teixeira de (2010), *Nova gramática do português brasileiro*, São Paulo, Editora Contexto.

Cavalcante, Sílvia/Duarte, Maria Eugênia Lamoglia (2008), *The subject position of infinitival sentences in Brazilian Portuguese: the embedding of a syntactic change*, Working Papers in Linguistics (Philadelphia, PA 14, 157–161.

Cedergren, Henrietta/Sankoff, David (1974), *Variable Rules: Performance as a statistical reflection of competence*, Language and Speech 50, 333–355.

Chomsky, Noam (1965), *Aspects of the Theory of Syntax*, Cambridge, MIT Press.

Chomsky, Noam (1981), *Lectures on government and binding*, Berlin/New York, Mouton de Gruyter.

Chomsky, Noam (1995), *The minimalist program*, Cambridge, MIT Press.

Coelho, Rafael F. (2006), *É nóis na fita! Duas variáveis linguísticas numa vizinhança da periferia paulistana. O pronome de primeira pessoa do plural e a marcação de plural no verbo,* Master's thesis, São Paulo, FFLCH-USP.

Colischonn, Gisela/Monaretto, Valéria de Oliveira (2012), *Banco de dados VARSUL: a relevância de suas características e a abrangência de seus resultados*, Alfa 56/3, 835–853.

Coseriu, Eugenio (1980), *"Historische Sprache" und "Dialekt"*, in: Joachim Göschel/Pavle Ivic/Kurt Kehr (edd.), *Dialekt und Dialektologie. Ergebnisse des internationalen Symposiums "Zur Theorie des Dialekts"*, Wiesbaden, Steiner, 106–115.

Duarte, Maria Eugênia Lamoglia (1986), *Variação e sintaxe: clítico acusativo, pronome lexical e categoria vazia no português do Brasil*, Master's thesis, São Paulo, Pontifícia Universidade Católica de São Paulo.

Duarte, Maria Eugênia Lamoglia (1989), *Clítico acusativo, pronome lexical e categoria vazia no português do Brasil*, in: Fernando Tarallo (ed.), *Fotografias sociolingüísticas*, Campinas, Editora da Unicamp, 49–50.

Duarte, Maria Eugênia Lamoglia (1993), *Do pronome nulo ao pronome pleno: a trajetória do sujeito no português brasileiro*, in: Ian Roberts/Mary A. Kato (edd.), *Português brasileiro. Uma viagem diacrônica*, Campinas, Editora da Unicamp, 107–128.

Duarte, Maria Eugênia Lamoglia (1995), *A perda do princípio "Evite Pronome" no português brasileiro*, Doctoral thesis, Campinas, UNICAMP.

Duarte, Maria Eugênia Lamoglia (2000), *The loss of the "avoid pronoun" principle in Brazilian Portuguese*, in: Mary A. Kato/Esmeralda Vailoti Negrão (edd.), *Brazilian Portuguese and the Null Subject Parameter*, Frankfurt, Vervuert, 17–36.

Eckert, Penelope (2000), *Linguistic Variation as Social Practice. The Linguistic Construction of Identity in Belten High*, Malden, Oxford, Blackwell Publishing.

Eckert, Penelope (2008), *Variation and the indexical field*, Journal of Sociolinguistics 12, 453–476.

Eckert, Penelope (2012), *Three waves of variation study: the emergence of meaning in the study of variation*, Annual Review of Anthropology 41, 87–100.

Ferguson, Charles A. (1959), *Diglossia*, Word 15, 325–340.

Ferrari, Lilian Ferreira (1994), *Variação lingüística e redes sociais no Morro dos Caboclos*, Doctoral thesis, Rio de Janeiro, Universidade Federal do Rio de Janeiro.

Ferrari, Lilian Ferreira (2008), *Reportar condicionais: uma questão de ponto de vista*, Revista de estudos de linguagem 16, 117–140.

Fraga, Letícia (2009), *Atitudes linguísticas e r-forte em Carambeí,* Acta Scientiarum/Language and Culture 31/2, 155–168.

Freitag, Raquel, et al. (2016), *Avaliação e variação linguística: estereótipos, marcadores e indicadores em uma comunidade escolar*, in: Raquel Freitag/Cristine Severo/Edair Görsky (edd.), *Sociolinguística e política linguística: olhares contemporâneos*, São Paulo, Blucher, 139–159.

Gonçalves, Sebastião Carlos Leite/Tenani, Luciani Ester (2008), *Problemas teórico-metodológicos na elaboração de um sistema de transcrição de dados interacionais: o caso do projeto ALIP (Amostra Lingüística do Interior Paulista),* Gragoatá 13/25, 165–183.

Guy, Gregory Riordan (1981), *Linguistic variation in Brazilian Portuguese. Aspects of the phonology, syntax, and language history,* Doctoral thesis, Philadelphia (PA), University of Pennsylvania.

Hora, Dermeval da (2004), *Estudos sociolingüísticos: perfil de uma comunidade*, Santa Maria, Pallotti.

Hora, Dermeval da (2009), *Vogais no ponto mais oriental das Américas*, João Pessoa, Editora Universitária.

Johnstone, Barbara (2004), *Place, Globalization, and Linguistic Variation*, in: Carmen Fought (ed.), *Sociolinguistic Variation. Critical Reflections*, New York, Oxford University Press, 65–83.

Koch, Peter/Oesterreicher, Wulf (1990), *Gesprochene Sprache in der Romania: Französisch, Italienisch, Spanisch*, Tübingen, Niemeyer. [Spanish version: *Lengua hablada en la Romania: español, francés, italiano*, Madrid, Gredos, 2007].

Krefeld, Thomas/Pustka, Elissa (edd.) (2010), *Perzeptive Varietätenlinguistik*, Frankfurt am Main u. a., Lang.

Kristiansen, Gitte/Dirven, René (edd.) (2008), *Cognitive sociolinguistics: Language variation, cultural models, social systems,* Berlin/New York, De Gruyter Mouton.

Labov, William (1966), *The Social Stratification of English in New York City*, Washington, D.C., Center for Applied Linguistics.

Labov, William (1972a), *The social motivation of a sound change*, in: William Labov (ed.), *Sociolinguistic Patterns*, Philadelphia, University of Pennsylvania Press, 1–42.

Labov, William (1972b), *Language in the Inner City*, Philadelphia, University of Pennsylvania Press.

Lambert, Wallace E., et al. (1960), *Evaluational reactions to spoken languages*, Journal of Abnormal and Social Psychology 60/1, 44–51.

Lemle, Miriam/Naro, Anthony J. (1977), *Competências básicas do português. Relatório final de pesquisa apresentado às instituições patrocinadoras*, Rio de Janeiro, Fundação MOBRAL/ Fundação Ford.

Lima, Izete de Souza/Lucena, Rubens Marques de (2013), *Influência de variáveis não linguísticas no processo de acomodação dialetal do /s/ em coda silábica por paraibanos em Recife*, Letrônica 6/ 1, 161–178.

Lope Blanch, Juan M. (1967), *Proyecto de estudio del habla culta de las principales ciudades de Hispanoamérica*, in: El Simposio de Bloomington, agosto de 1964: Actas, informes y comunicaciones, Bogotá, Instituto Caro y Cuervo, 255–266.

Lopes, Célia Regina dos Santos (1998), *"Nós" e "a gente" no português falado culto do Brasil*, D.E.L.T.A. 14/2, 405–422.

Lopes, Célia Regina dos Santos (2003), *A inserção de "a gente" no quadro pronominal do português*, Madrid/Frankfurt a.M., Iberoamericana/Vervuert.

Lopes, Célia Regina dos Santos (2007), *A gramaticalização de "a gente" em português em tempo real e de curta duração: retenção e mudança na especificação dos traços intrínsecos*, Forum Linguístico 4, 47–80.

Lopes, Célia Regina dos Santos/Callou, Dinah (2004), *Contribuições sociolinguísticas para o ensino e a pesquisa: a questão da variação e mudança linguística*, Revista do GELNE 5, 63–74.

Lucchesi, Dante (2001), *As duas grandes vertentes da história sociolingüística do Brasil*, D.E.L.T.A. 17/ 1, 97–132.

Mello, Antonio Augusto Souza (1996), *Genetic affiliation of the language of the Indians Aure and Aura*, Opción 12/19, 67–81.

Mendes, Ronald Beline (2015), *Diminutives and Masculinity in Brazilian Portuguese*, in: Tommaso Milani (ed.), *Language and Masculinities. Performances, Intersections, Dislocations*, New York, Routledge, 117–132.

Mendes, Ronald Beline (2016), *Nonstandard plural noun phrase agreement as an index of masculinity*, in: Erez Levon/Ronald Beline Mendes (edd.), *Language, sexuality and power: studies in intersectional sociolinguistics*, Oxford, Oxford University Press, 105–129.

Mendes, Ronald Beline (2018), *Percepção e performance de masculinidades. Efeitos da concordância nominal de número e da pronúncia de /e/ nasal*, Habilitation thesis, Universidade de São Paulo.

Mendes, Ronald Beline/Oushiro, Livia (2012), *O paulistano no mapa sociolinguístico brasileiro*, Alfa 56/3, 973–1001.

Naro, Anthony J. (1981), *The social and structural dimensions of a syntactic change*, Language and Cognition 57/1, 62–98.

Nascentes, Antenor (1922), *O linguajar carioca em 1922*, Rio, Süssekind de Mendonça.

Ochs, Elinor (1992), *Indexing gender*, in: Alessandro Duranti/Charles Goodwin (edd.), *Rethinking context. Language as an interactive phenomenon*, Cambridge, Cambridge University Press, 335–358.

Oushiro, Livia (2015), *Identidade na pluralidade. Avaliação, produção e percepção linguística na cidade de São Paulo*, Doctoral thesis, São Paulo, Universidade de São Paulo.

Pagotto, Emilio Gozze (2004), *Variação e(é) Identidade*, Maceió, ED/UFAL/EDUFBA.

Paiva, Maria/Scherre, Marta (1999), *Retrospectiva sociolinguística: contribuições do PEUL*: Documentação de Estudos em Lingüística Teórica e Aplicada 15, 201–232.

Pinto, Edith Pimentel (1990), *A Gramatiquinha de Mário de Andrade: texto e contexto*, São Paulo, Duas Cidades.

Ploog, Katja/Reich, Uli (2005), *Rasgos socio-indexicales en la dinámica urbana*, Lexis 29/1, 47–78.

Ploog, Katja/Reich, Uli (2006), *Urbane Prozesse. Migration und Variation in Lima, São Paulo und Abidjan*, in: Thomas Krefeld (ed.), *Modellando lo spazio in linguistica/Modellierung sprachlicher Räume*, Frankfurt, Lang, 215–256.

Podesva, Robert J. (2006), *Phonetic detail in sociolinguistic variation. Its linguistic significance and role in the construction of social meaning*, Doctoral thesis, Stanford (CA), Stanford University.

Preti, Dino/Urbano, Hudinilson (1990), *A linguagem falada culta na cidade de São Paulo*, vol. 4: *Estudos*, São Paulo, Ta Queiroz.

Reich, Uli (2002), *Freie Pronomina, Verbalklitika und Nullobjekte im Spielraum diskursiver Variation des Portugiesischen in São Paulo*, Tübingen, Narr.

Reich, Uli (2009), *Mille plateaux linguistiques*, Neue Romania 39, 287–299.

Ribeiro, Branca Telles/ Garcez, Pedro M. (edd.), (1998), *Sociolingüística interacional: antropologia, linguística e sociologia em análise do discurso*, Porto Alegre, AGE.

Rodrigues, Angela C. Souza (1987), *A concordância verbal no português popular em São Paulo*, Doctoral thesis, São Paulo, Universidade de São Paulo.

Rodrigues, Angela C. Souza (2000), *Concordância verbal e saliência social no português popular em São Paulo*, in: Eberhard Gärtner/Christine Hundt/Axel Schönberger (edd.), *Estudos de sociolingüística brasileira e portuguesa*, Frankfurt, TFM, 41–61.

Rodrigues, Angela Cecília de Souza (2004), *Concordância verbal, sociolingüística e história do português brasileiro*, Fórum Linguístico 4, 115–145.

Salomão, Ana Cristina Biondo (2011), *Vizinhança global ou proximidade imposta? Impactos da comunicação intercultural mediada por computador sobre o papel da cultura no ensino de língua inglesa*, DELTA: Documentação e Estudos em Linguística Teórica e Aplicada 27/2, 235–256.

Scherre, Maria Marta Pereira (1978), *A regra de concordância de número no sintagma nominal em português*, Master's thesis, Rio de Janeiro, Universidade Federal do Rio de Janeiro.

Scherre, Maria Marta Pereira (1981). *La variation de la règle d'accord du nombre dans le syntagme nominal en Portugais*, Variation Omnibus 212/6, 125–133.

Scherre, Maria Marta Pereira (1996), *Sobre a influência de variáveis sociais na concordância nominal em português*, in: Giselle Machline de Oliveira e Silva/ Maria Marta Pereira Scherre. (eds.),

Padrões sociolingüísticos—Análisis de fonômenos variáveis do português falado no Rio de Janeiro, Rio de Janeiro, Tempo Brasileiro, 239–264.

Scherre, Maria Marta P./Naro, Anthony Julius (1998), *Sobre a concordância de número no português falado do Brasil*, in: Giovanni Ruffino (ed.), *Atti del XXI Congresso Internazionale di Linguistica e Filologia Romanza*, vol. 5: *Dialettologia, geolinguistica, sociolinguistica*, Tübingen, Niemeyer, 509–523.

Silva, Giselle Machline de Oliveira/Scherre, Maria Marta Pereira (edd.) (1996), *Padrões sociolingüísticos: análise de fenômenos variáveis do português falado na cidade do Rio de Janeiro*, Rio de Janeiro, Tempo Brasileiro.

Silva, Hélen Cristina da/Aguilera, Vanderci Andrade de (2014), *O poder de uma diferença. Um estudo de crenças e atitudes linguísticas*, Alfa 58/3, 703–723.

Silverstein, Michael (2003), *Indexical order and the dialectics of sociolinguistic life*, Language and Communication 23, 193–229.

Tarallo, Fernando (1983), *Relativization strategies in Brazilian Portuguese*, Doctoral thesis, Philadelphia (PA), University of Pennsylvania.

Tarallo, Fernando (ed.) (1989), *Fotografias sociolingüísticas*, Campinas, Editora da Unicamp.

Tarallo, Fernando/Kato, Mary (1989), *Harmonia trans-sistêmica: variação intra- e inter-lingüística*, Preedição 5, 1–41.

Tarallo, Fernando/Kato, Mary (2007), *Harmonia trans-sistêmica: variação intra- e inter-lingüística*, Diadorim 2, 13–42.

Vitral, Lorenzo (1996), *A forma CÊ e a noção de gramaticalização*, Revista de estudos linguísticos 4/1, 115–124.

Weinreich, Uriel/Labov, William/Herzog, Marvin (1968), *Empirical foundations for a theory of language change*, in: Winfried Lehmann/Yakov Malkiel (edd.), *Directions for Historical Linguistics,* Austin, University of Texas Press, 95–195

Zilles, Ana (2005), *The development of a new pronoun. The linguistic and social embedding of "a gente" in Brazilian Portuguese*, Language Variation and Change 17, 19–53.

John M. Lipski

11 Brazilian Portuguese: contemporary language contacts

Abstract: Brazilian Portuguese has been enriched by many languages arriving as the result of post-colonial immigration. Among the most common are German varieties (Hunsrückisch and Pomeranian) and northern Italian Veneto varieties (Talian). Also present in Brazil are bilingual communities using Ukrainian, Polish, Dutch, Chinese, Korean, Japanese, Haitian Creole, English, Spanish, and other languages deriving from immigration. These languages are most tenaciously retained in rural agricultural communities or *colônias*, where effects on local varieties of Portuguese can frequently be observed. In urban areas of Brazil, most immigrant languages have given way to Portuguese monolingualism, except for a few neighborhoods harboring large numbers of recent immigrants, e.g., in São Paulo. Brazilian Portuguese is also spoken in northern Uruguay and in Misiones province, Argentina, in contact with Spanish, and partially hybridized *Portunhol* varieties have emerged. Brazilians in the diaspora maintain Portuguese in bilingual contexts in the United States, Japan, and Paraguay.

Keywords: Brazilian Portuguese, German, Italian, Talian, Hunsrückisch, Pomeranian, Ukrainian, Polish, Chinese, Korean, Japanese, Spanish, English, Haitian Creole

1 Introduction

Brazil, the largest nation in South America, has been the recipient of many waves of immigration, particularly in the post-colonial period, and several languages resulting from 19th and 20th-century immigration continue to be spoken in contact with Portuguese. In most instances these languages are confined to ethnic communities of foreign origin, which have remained in substantially rural areas of the southern Brazilian states. Taken together, however, the number of Brazilians who speak more than one language is not inconsiderable, and has contributed to the mosaic of macro- and micro-dialectal variation in Brazilian Portuguese. This chapter will concentrate on post-colonial contacts (see chapter 3 for the history of contacts, including indigenous and African contacts).

Although immigration to Brazil has occurred since the country's independence in 1822, massive (voluntary) immigration of speakers of languages other than Portuguese only began in the final decades of the 19th century, and reached a peak during the 20th century, with some six million immigrants arriving during this time span. Immigration from Portugal was always significant, often representing the largest proportion of foreign arrivals, but several other central and southern European languages also entered

https://doi.org/10.1515/9783110405958-012

the mix, later joined by Asian languages. The first large wave of 19th century non-Portuguese immigration included speakers of various German dialects, later followed by Italian speakers, and then by speakers of Slavic languages, especially Ukrainian and Polish. Immigration from Japan and China, and to a lesser extent Korea, occurred principally during the 20th century, but on a reduced scale continues to the present. Spain also contributed heavily to the immigrant population in Brazil, but few descendants of Spaniards maintained the Spanish language past the first generation. Much of the Spanish immigration came from Galicia, whose language is even closer to Portuguese, further easing the linguistic transition.

With the exception of occasional lexical borrowings, in general regionally distributed, these languages have not made a national impact on Brazilian Portuguese. Within the bilingual speech communities, however, contact-induced traits can be found in local varieties of Portuguese, at times extending to heritage speakers of the immigrant languages and even monolingual Portuguese speakers in the same communities. Naturally, Portuguese influence on the immigrant languages is more frequent, especially in subsequent generations of heritage speakers, and forms the basis for much of the bibliography on bilingualism in Brazil. A complete survey of contemporary immigrant-derived language-contact environments in Brazil would exceed the scope of this article, as would an account of the dozens of indigenous languages currently in contact with Brazilian Portuguese. In the following sections attention will be devoted to a selection of the largest and most prominent immigration-derived bilingual speech communities in Brazil; whenever possible, the effects of language contact on regional varieties of Portuguese will be included. The discussion also includes varieties of Brazilian Portuguese outside the borders of Brazil.

2 Germanic languages and dialects in contact with Portuguese in Brazil

After Portuguese, German (in dialectal varieties) is the language with the greatest number of speakers in Brazil. The Ethnologue data base estimates a (perhaps unrealistically high) total of at least three million speakers of Brazilian *Hunsrückisch* and another million and a half speakers of other German varieties, most prominently Pomeranian. The German immigrant population is concentrated in the southern states of Santa Catarina, Espírito Santo, and Rio Grande do Sul (Maltzahn 2013; Rost 2008). São Paulo state also contains many German-speaking communities (Silva 2010). The German dialects spoken in the southern Brazilian states have been studied extensively (e.g., Altenhofen 199; Meyer 2009; Fritzen 2007; Rosenberg 1998; 2002; 2003a; 2003b; 2012; Schaumloeffel 2003), including code-switching and contact with Portuguese. German immigration to Brazil began in the early decades of the 19th century and reached its peak in the 20th century between the two world wars. During most of

the period of German immigration the settlers lived in German-speaking colonies, some of which are still in existence. Since much of the immigration occurred before the consolidation of modern Germany and the general acceptance of a High German (*Hochdeutsch*) standard, a number of regional Germanic dialects are found in Brazil; the two most prevalent are *Hunsrückisch*, from the Rhineland Palatinate area (Altenhofen 1996; Altenhofen/Frey 2006; Barros/Philippsen 2013; Damke 1997; Leão 2007; Meyer 2009), and *Pomersch* or Pomeranian (Beilke 2013a; 2013b; Blank 2013; Bremenkamp 2014; Höhmann/Savedra 2011; Schaeffer 2012; Schneider 2012; Vandresen/Rodrigues Corrêa 2008), from a region now part of Poland. Although Pomeranian has all but disappeared from Europe and the Palatinate dialects are on the wane, both are vigorously maintained in some southern Brazilian cities. Also found in Brazil is Westfalian (Luersen 2009). As in some other regions of Latin America, Mennonite Low German (Plautdietsch) is still spoken in some Brazilian communities (Dück 2011, 2012), often as the dominant first language.

Many Brazilian German speakers, including some linguists, have identified a Portuguese-influenced Germanic koine popularly referred to as *Brasildeutsch* (e.g., Damke 2008; Souza 2002; Putnam/Lipski 2016 for nearby Misiones province, Argentina), as well as Germanic influence on local varieties of Portuguese (e.g., Vandresen 2006). Most such influence is confined to the domain of phonetics; Bandeira/Zimmer (2012) and Zimmer/Bandeira (2009) study the voice onset timing (VOT) of Pomeranian-Portuguese bilinguals, which are typically longer than those of Portuguese monolinguals. Lara and Battisti (2014) provide similar data on the VOT of Hunsrückish-Portuguese bilinguals. Benincá (2009) and Rodrigues (2009) provide additional observations of Pomeranian-Portuguese bilingualism, including voiceless articulation of word-initial voiced obstruents in Portuguese.

The Portuguese of recent German-speaking arrivals is studied by Rezagli (2010), e. g. *para* + INFINITIVE constructions in relation to German *zu* + INFINITIVE. In some communities, heritage German and Italian varieties are in contact (Krug 2004; Sufredini 1993), with as yet unexplored linguistic characteristics.

Several Brazilian communities have adopted German varieties as co-official languages; for Hunsrückisch this includes some cities in Santa Catarina and Rio Grande do Sul, and for Pomeranian some localities in Espírito Santo, Minas Gerais, Santa Catarina, and Rio Grande do Sul. The states of Espírito Santo and Rio Grande do Sul officially acknowledge German varieties as part of the states' linguistic heritage.

The first contact between Portuguese and Dutch occurred during the period 1630–1654, when the Dutch occupied the northeastern portion of the then Portuguese colony, and the state of Pernambuco remained officially part of the "Dutch Republic" until 1661. These early contacts left few traces on the language and it was not until the late 19th and early 20th century, and especially right after World War II that substantial numbers of Dutch settlers formed communities in Brazil where the Dutch language evolved in contact with Portuguese. Bilingual Dutch-Portuguese communities are still found in the state of Paraná (Fraga 2009a; 2011; Rickli 2003), and purported Dutch in-

fluence on local Portuguese includes use of the alveolar trill /r/ instead of the velar fricative [x] more common in Brazilian Portuguese (Fraga 2009b).

3 Italian dialects in contact with Portuguese in Brazil

Italian immigration to Brazil was massive, particularly in the final decades of the 19th century, when more than one million Italians settled, mainly in the southern region (Cenni 2003). The Italian population was most concentrated in São Paulo state (where Italian varieties have influenced the vernacular Portuguese of the city of São Paulo: Alves/Maroneze 2004; Caprara/Mordente 2004; Castro 1997), followed by Rio Grande do Sul and Minas Gerais. Subsequent internal migration resulted in significant Italian communities in the neighboring states of Santa Catarina and Paraná. The greatest proportion of Italian immigration came from the northern region, especially during the first decades, with the Veneto-speaking zone topping the list. The Veneto language continues to be spoken in some areas of Rio Grande do Sul and Santa Catarina (Dal Picol 2013; Peres 2011), and was once prominent in São Paulo as well (Vieira 2010). Veneto and other northern Italian dialects, e.g., Trentino (Boso 1991; Schissatti 2014) are often known as *Talian* in Brazil (Armilato 2012; Pertile 2009. The municipality of Serafina Corrêa (Rio Grande do Sul) has adopted *Talian* as a co-official language, and the states of Rio Grande do Sul and Santa Catarina officially acknowledge *Talian* as integral components of the states' linguistic heritage. Estimates of the current number of *Talian* speakers range from 500,000 to more than three million, with the latter figure undoubtedly including heritage speakers with less than total fluency. The northern Italian Tirolean dialect is also spoken in some communities (Leopoldino 2009; 2014).

Numerous regional vernacular Portuguese traits have variously been attributed to contact with Veneto and other Italian regional dialects (overviews are given in Faggion 2010; Frosi/Mioranza 1983; Margotti 2004; Ponso 2003; Toscan 2005). One example involves the assignment of grammatical gender to nouns, e.g., the canonical *a alface* 'lettuce' (f.) vs. *o alface* (m.) as used by some Italian-Portuguese bilingual speakers (Borges 2007). Another is raising of mid vowels to harmonize with the diminutive suffixes *-inho* and *-zinho* (Faggion 2006), for example *escola* 'school' > *isculinha*. Contact with Italian is also implicated in other vernacular traits of southern Brazilian Portuguese, including non-palatalization of /t/ and /d/ before [i], non-raising of atonic final /e/ and /o/ as well as lowering of final atonic /u/ as in *ele caiu* 'he fell'), lack of velarization of /l/ in syllabic codas (Espiga 2003), realization of the nasal diphthong *-ão* as [õ] (Faggion 2010; Tomiello 2005), unreduced (not centralized) realization of nasal /ã/ and final atonic /a/, presence of front rounded vowels [y] and [ø], pluralization with *-i* instead of *-s* (Leopoldino 2009, 2014), and neutralization of single /ɾ/ and trill /r/ in favor of the former (Bovo 2004; Brescancini/Monaretto 2008; Frosi 1987; Frosi et al. 2005; 2007; 2008; Loriato/Ponzo Peres 2013).

4 Slavic languages in contact with Portuguese in Brazil

Brazil received considerable immigration of Ukrainian speakers, particularly to the state of Paraná, and is home to the third largest group of Ukrainians in the Americas (more than 500,000). Immigration began towards the end of the 19th century, accelerated following World War I, and reached its peak in the aftermath of World War II (Boruszenko 1981; Garcia 2009; Lemke 2010). The first generations maintained the Ukrainian language and instituted classes in Ukrainian, but shift to Portuguese has generally occurred in urban areas (Mezavila 2008), while Ukrainian continues to be spoken in many of the rural *colônias* of Paraná. Garcia (2009) provides sociophonetic data on the realization of the Portuguese nasal diphthong *-ão* among rural Ukrainian-Portuguese bilinguals, where older bilinguals use the variants [õw̃] and [ɔ̃w̃] to a greater extent than Portuguese monolinguals. Lemke (2010) analyzes the acquisition of Spanish by Portuguese-Ukrainian bilinguals; only Portuguese influence in their Spanish is evident.

Polish immigration to Brazil followed similar patterns to those of Ukrainians, beginning in the late 19th century but intensifying after 1920. Currently there are some 1.8 million Polish-Brazilians, mainly concentrated in the state of Paraná, where more than 700,000 are found. The Polish language has not been maintained as extensively as Ukrainian, but continues to survive in some villages and rural areas near Curitiba (Dvorak 2013; Ferraz 2007; Malczewski 2010; Miodunka 2003). There is little documentation on Polish influence on local varieties of Portuguese (e.g., Druszcz 1984 for subtle phonetic and morphosyntactic traits such as variable word order, inconsistent gender and number agreement, lack of determiners, neutralization of /ɾ/ and /r/, over-differentiation of the Portuguese vowel system). Portuguese influence has also been detected among heritage Polish speakers (Costa/Gielinski 2014).

5 Japanese in contact with Portuguese in Brazil

The largest population of Japanese origin outside of Japan is found in Brazil, with numbers approaching two million, of which an estimated 380,000 still speak at least some Japanese (ethnologue.com). The Japanese immigration to Brazil began in 1908 and reached its peak in the period 1910–1940, with another spike occurring in the late 1950's. Japanese immigrants were often viewed with ambivalence in Brazil, which together with Japanese cultural pride resulted in the tenacious retention of the Japanese language beyond the first generation (Adachi 1997, 2001; Kanazawa and Loveday 1988). Brazilian Portuguese has definitively influenced the Japanese spoken in Brazil, while Brazilian Portuguese shows little influence of Japanese, beyond the L2 approximations to Portuguese produced by first-generation Japanese immigrants (Kanazawa

and Loveday 1988: 430; Fuchs 1996). Lingering traces of Portuguese interference can be found in some Brazilian-Japanese communities (e.g., Fujiwara 2014; Gibo 2014; Ota 2009; Saiki 2013). Parlato-Oliveira et al. (2010) demonstrate that Japanese-Brazilian Portuguese bilinguals perceive illusory or "ghost" vowels /i/ or /u/ in non-permissible phonotactic sequences, in accordance with their degree of bilingualism.

6 Chinese and Korean in contact with Portuguese in Brazil

Brazil has a considerable population of Chinese origin (estimated at around 250,000), stemming from immigration that began in the early 19th century and which continues at the present time (Jye et al. 2006). Immigration spiked during World War II and after the creation of the People's Republic of China in 1949. Other upsurges came from Taiwan following the United Nations' recognition of mainland China in 1971, and from Hong Kong following the incorporation into China in 1997. With the exception of the Taiwanese, most Chinese immigrants to Brazil spoke Cantonese and neighboring Chinese languages. Those arriving from Macau and some from Hong Kong also spoke Macau Portuguese creole and even some non-creole Portuguese. Although many Chinese immigrants strove to maintain their language and transmit it to future generations, language displacement in favor of Portuguese has been the usual outcome, with Chinese-influenced Portuguese typically characterizing only the first generation of Chinese-Portuguese bilinguals. More recent Chinese immigration is centered around Chinese-owned businesses, especially in São Paulo (Piza 2012), but there is as yet little information on the Portuguese acquired by the newest Chinese arrivals.

Beginning in the 1960's many Koreans immigrated to Brazil, particularly to the state of São Paulo, where they experienced some difficulty in achieving social and linguistic integration (Choi 1991; 1998; Valim et al. 2011). At present there may be some 37,000 Korean speakers in Brazil (ethnologue.com), out of a total of more than 50,000 Korean-Brazilians. While little is known of these Korean's use of Portuguese, general tendencies can be extrapolated from the study of Koreans' L2 Portuguese in other settings (e.g., Kim 2005).

7 Creole languages in Brazil

Brazil's northern border impinges on Surinam, Guyana, and French Guiana, nations in which one or more Afro-European creole languages form part of the linguistic landscape. Particularly near the border with French Guiana, some French Creole is spoken in Brazil (Day 2013; 2017; Silva 2017). The endangered language known as Karipúna French Creole or Lanc Patuá, apparently related to French Guiana Creole, still survives

among the Karipúna indigenous community in the state of Amapá (Andrade 1984; Tobler 1983; 1987). Along the border with Guyana (e.g., in Bonfim, Roraima), some Guayanese Creole English is present, together with non-creole English (Machado/Pereira 2019; Prudente 2019; Santos 2012). Along the border with Suriname, only indigenous languages but none of the Surinamese creole languages (Srnan Tongo, Saramaccan, Ndjuka, etc.) are spoken.

One of the most dramatic examples of heritage language enclaves in the Americas results from the Haitian diaspora, an ongoing demographic movement propelled by unfavorable economic and political conditions in Haiti. In Brazil as in other Latin American nations, Haitian immigrants tend to live together, which reinforces the maintenance of Haitian *Kreyòl* past the first generation (Alessi 2013; Cogo 2014; Cotinguiba/Cotinguiba 2014a; 2014b; Godoy 2011; Moraes 2013; Nieto 2014). Social discrimination and racism, frequently combined with undocumented immigrant status, further isolates Haitians, and also contributes to heritage language survival.

Fig. 1: Spanish-speaking border communities.

8 United States English: the "Confederados"

Following the United States Civil War (1861–1865), some 20,000 individuals from the defeated secessionist southern states (especially Georgia, Alabama, and Texas) emigrated to Brazil, welcomed by the Brazilian emperor Dom Pedro II, who wanted to expand cotton production in Brazil. Many of the descendants of these "Confederados," living in the communities of Americana and Santa Bárbara d'Oeste near São Paulo, still speak a southern variety of American English, increasingly influenced by Portuguese through borrowing and code-switching (Lownes 2018; Medeiros 1982; Montgomery/Melo 1990; 1995; Pérez Gómez 2018; Weaver 1961). Some of these descendants of southern United States immigrants have visited the United States and many maintain correspondence with post-Confederacy organizations in the southern United States. In 1972, while still governor of Georgia, the former U. S. president Jimmy Carter visited the community and spoke English with many of the residents. Within Brazil there is an American Descendents' Association (*Associação Descendência Americana*) and an annual Confederadaos festival.

9 Within, along and across the border: Brazilian Portuguese in contact with Spanish

The third-largest group of immigrants to Brazil (after Italy and Portugal) came from Spain. Many were from Galicia, and their Galician language permitted an easy transition to Portuguese. Spanish speakers similarly acquired Portuguese, and the number of heritage Spanish-speaking descendants of immigrants from Spain has never been large. At present, Spanish speakers in Brazil principally represent immigration from neighboring nations. As South American's largest nation, Brazil shares a border with every Spanish-speaking nation except Ecuador and Chile, and along the border many Portuguese-Spanish contact phenomena can be observed. There are a number of border-straddling communities in which Portuguese is used as a second or sometimes first language within the respective Spanish-speaking countries, but in the majority of instances the language-contact configurations are asymmetrical:

Just beyond Brazil's borders Portuguese is frequently spoken in neighboring areas of nominally Spanish-speaking countries. In most Spanish-speaking border communities Portuguese is only used when speaking to Brazilians, although passive competence is often high due to the predominance of Brazilian radio and television. In addition to the many border regions where native Spanish speakers learn Portuguese as a second language to speak to Brazilians, there are two dialect zones where vernacular Portuguese—and possible also Portuguese-Spanish hybrid varieties—are spoken natively, and where stable language-contact phenomena can be observed. The first is the widely studied northern Uruguayan region. Another important dialect cluster, as yet

little studied and with unique sociolinguistic features not found in northern Uruguay, is located in the extreme northeastern Argentine province of Misiones, where Portuguese is spoken as a first language in several rural communities and small towns (Lipski 2015).

In the following Spanish-speaking border communities (shown on Map 1) Portuguese is spoken as a second language predominantly with Brazilians, for commercial purposes, and at least some Portuguese influence can often be found in the local Spanish dialects (Lipski 2006; 2008; 2009; 2010; 2011a; 2011b; 2015; Kaufmann 2009): Paso de los Libres (Argentina), Cobija, Guayaramerín, and Bolpebra (Bolivia), Leticia (Colombia), Capitán Bado, Pedro Juan Caballero, Bella Vista Norte, and Zanja Pytã (Paraguay), Iñapari and Santa Rosa (Peru), and Santa Elena de Uairén (Venezuela). In these nominally Spanish-speaking communities near the Brazilian border Spanish is mixed into Portuguese, usually unconsciously and frequently involuntarily. L2 Portuguese as produced by Spanish speakers in border regions differs from natively spoken Portuguese in Uruguay and Misiones as well as from neighboring Brazilian varieties. For example, in addition to having a high density of Spanish-Portuguese juxtaposition, hybrid combinations produced spontaneously by L2 Portuguese speakers in these border communities appear to violate well-documented syntactic constraints on intrasentential code-switching.[1] These include putative Spanish-Portuguese switching between subject pronouns and verbs, after fronted interrogative words, between negators and verbs, and between auxiliary verbs and main verbs. The bilingual incursions are not the result of community-wide traditions of language shifting, since neither the first- nor second-language Portuguese speakers engage in the type of fluid and self-aware code-switching that is characteristic of many bilingual groups. Rather, the near-absence of sociolinguistic strictures and minimal formal training in one or both languages provide environments in which the natural intertwining of these closely related languages can take place.

Within Brazil Spanish is not typically spoken in border communities and Brazilians tend to use Portuguese when traveling to neighboring Spanish-speaking communities. One exception to this trend is Chui in extreme southeastern Rio Grande do Sul state, where Spanish is used as frequently as in the homophonous Uruguayan sister city Chuy; the two cities share an unrestricted land border marked by a single street (Couto 2011; Lipski 2015). Espiga (1997; 2002; 2003; 2006) attributes to Spanish influence the lack of vocalization to [w] or velarization of coda /l/ along Brazil's southern border with Uruguay.

In the interior of Brazil, Spanish-speaking immigrant groups from neighboring countries have settled in urban areas, particularly São Paulo. The largest enclaves are from Bolivia, predominantly Aymara speakers from the highlands, and who live in

1 The putative syntactic constraints on intrasentential code-switching discussed here are reviewed in Lipski (1985), Muysken (2000), and many of the studies in Bullock/Toribio (2009).

marginal conditions in the city's poorest neighborhoods (Cymbalista 2007; Silva 2005; 2006; Simai/Baeninger 2012; Souchaud 2010; Souchaud/Baeninger 2008). Most Bolivians work in construction or the textile industry, and many lack legal immigrant status and are especially vulnerable. At present there may be more than 200,000 Bolivians in São Paulo (Niehoff 2011, 29), and an undetermined number in other Brazilian cities. Niehoff (2011) has documented the L2 Portuguese of Bolivian immigrants in São Paulo, which exhibits the usual range of Spanish interference phenomena. At the segmental level, Bolivians have difficulty with the Portuguese seven-vowel system, nasal vowels and diphthongs, the voiced phoneme /z/ and the syntactically-motivated voicing of intervocalic /s/ across word boundaries, vocalization of coda /l/, palatalization of /t/ and /d/ before /i/, and realization of the "trill" /r/. At the morphosyntactic level, Bolivians exhibit many errors of grammatical gender, not only with cognate words whose gender differs between Spanish and Portuguese (e.g., *sangue* 'blood,' *leite* 'milk') but also with items that share the same gender in both languages. Bolivians have adopted the vernacular Brazilian Portuguese tendency to mark plural /s/ only on the first element of noun phrases, as well as the gravitation towards the third-person singular as invariant verb form. Unlike native Portuguese speakers, Bolivians sometimes employ the 3s form for first-person singular as well. There is also some (probably unintentional) code-switching, including non-constituent switching best characterized as congruent lexicalization as defined by Muysken (2000, chap. 5); Niehoff (2011, 66, 123–125, 173–176) offers examples.

More recently, the economic collapse of Venezuela has driven thousands of Venezuelans across the southern border into Brazil, where nuclei of Venezuelan Spanish speakers have already emerged (Almeida/Santi 2018; Pimentel/Cotinguiba 2014; Santos/Vasconcelos 2016; Simões 2017).

10 Just beyond the border: vernacular Portuguese of Misiones, Argentina

The province of Misiones, in the extreme northeast of Argentina, is a narrow peninsula bordered by Paraguay to the west and Brazil to the east. Throughout the eastern portion of the province Portuguese continues to be the dominant language in all rural regions and in some urban centers as well. Portuguese is spoken as a home language in rural areas of Misiones from Puerto Iguazú and Comandante Andresito in the north all along the eastern corridor near the Uruguay River as far south as San Javier. The greatest concentration of speakers is found between the river town of El Soberbio and the village of Colonia Aurora with its rural dependencies Colonia Alicia and Puerto Londero. El Soberbio stands out among all the communities in Misiones in that vernacular Portuguese is the principal language of the entire community, spoken in businesses, government offices, and in formal and informal exchanges. The Portuguese spoken in

Misiones bears the mark of rural vernacular Brazilian Portuguese of the neighboring states, and has little resemblance to urban standardized Portuguese. Basic descriptions can be found in Lipski (2011b; 2015; 2018a; 2018b; 2019; 2020), Maia (2004), and Maia/Triches (2017); Misiones Portuguese is also mentioned by Daviña (2003) and Sturza (1994). Despite the fact that most residents of the *chacras* (small farms) of eastern Misiones listen to Brazilian Portuguese media, there is no attempt to emulate prestigious Brazilian varieties, unlike the situation in northern Uruguay (Carvalho 2004a). In most of these communities at least some Portuguese is also spoken as a second language by individuals involved in commerce with Brazil or with visiting Brazilians, but only in a few communities is L2 Portuguese usage extensive (e.g., in Bernardo de Irigoyen, which has an open land border with Brazil). The Portuguese of Misiones has relevance for the speech of adjacent Brazilian communities since many Portuguese speakers in Misiones have family ties to Brazil and frequently visit that country and interact using Misiones Portuguese.

As in northern Uruguay, the massive presence of Spanish in Misiones arrived with immigrants from more populated regions of the country. The principal difference is that in Misiones the use of Portuguese never provoked the harsh rejection found in northern Uruguay. Misioneros' tolerance of "Portuñol" is much greater than non-Portuguese-speaking Uruguayans' views towards *fronterizo* dialects. Although most of the schools in the Portuguese-speaking communities in Misiones do not offer classes in Portuguese, and despite the fact that children from Portuguese-speaking households frequently experience language difficulties in the first years of school, there has never been a campaign to eradicate Portuguese, or to label rural vernacular speech with terms like *rompe-idioma* 'language-breaker' as used in Uruguay. With the exception of towns such as El Soberbio and Santa Rita, Portuguese speakers in Misiones are generally rural residents, but speaking Portuguese is not publicly criticized or explicitly correlated with working-class or rustic speech. Misioneros' views toward Portuguese often involve gently humorous acknowledgement of "portuñol." In Misiones, the fact that Portuguese is most frequently spoken in rural regions contributes to its tenacity. Unlike in urban areas, where interaction with Spanish speakers inevitably results in language mixing and ultimately in the displacement of Portuguese by Spanish, on the small farms of Misiones residents spend most days speaking to one another only in Portuguese. This fact, combined with the relatively little sociolinguistic stigmatization of Portuguese/"portunhol", results in Misiones Portuguese having comparatively fewer Spanish incursions than Uruguayan *fronterizo*.

Although there are numerous language-mixing phenomena in the aforementioned Misiones communities, what is commonly referred to as "portunhol" is grounded in the vernacular varieties of Brazilian Portuguese from the neighboring states (e.g., Meirelles 2007), combined with some regional Spanish traits. The most robust features of Misiones Portuguese, all found to some extent in neighboring varieties of vernacular Brazilian Portuguese, are:

(1) The alveolar pronunciation [t], [d] of /t/ and /d/ before [i], instead of the pre-palatal affricates [ʧ]/[ʤ] found in Rio de Janeiro, São Paulo, and other major urban dialects in Brazil: *cidade* [si.'da.di] 'city.' This same pronunciation is also found in northern Uruguayan "portuñol," although imitation of the more prestigious São Paulo-oriented palatalized pronunciation is gaining ground (Carvalho 2004a).
(2) The delateralized pronunciation [j] of the palatal lateral phoneme /ʎ/: *trabalhamos* [tra.ba.'ja.mo], *mulher* [mu.'je]. This tendency is also found in northern Uruguay and is common—although stigmatized—throughout much of Brazil.
(3) Loss of final /ɾ/ in verbal infinitives: *trabalhar* [tra.ba.'ja], *estudar* [es.tu.'da]. This feature also characterizes northern Uruguay and many regional vernacular varieties within Brazil.
(4) Alveolar lateral pronunciation [l] of posnuclear /l/ (*Brasil* [bra.'zil]) in alternation with the vocalized pronunciation [w] found in most Brazilian dialects. The alveolar lateral is also found in northern Uruguay, in alternation with the vocalized pronunciation. It is relatively uncommon in Brazil, and may be due to Spanish influence.
(5) Realization of /r/ as an alveolar trill (*carro* ['ka.ro]) instead of the velar fricative [x] found in many regions of Brazil. This feature is also found in northern Uruguay, alternating with the velar fricative pronunciation. The alveolar trill is found in many southern Brazilian Portuguese dialects.
(6) Use of the second person singular subject pronoun *tu* (+ third person singular verb forms) instead of the more usual Brazilian *você*, and the possessive *tu* instead of canonical Portuguese *seu, sua, teu, tua*: *tu mora com tu família* 'you live with your family.' This feature is also found in northern Uruguay and in much of southern Brazil (with and without 2nd person singular verbal agreement).
(7) "Stripped" plural noun phrases, with /s/ marked only on the first element, usually a determiner: *os irmão* 'the siblings,' *as pessoa* 'the people.' This feature is also found in northern Uruguay and throughout Brazil, but in highly variable fashion. Within Brazil the rate of depleted plural marking appears to be decreasing in recent generations, due to greater access to public education as well as perhaps increased penetration of communications media (Guy 1981, 452–453). For example, in Rio de Janeiro in the 1970's the rate of omission of plural /-s/ in the second position of noun phrases was 72 % whereas a cross-section including more educated speakers from São Paulo collected around 2010 showed a rate of only around 14 %. Vernacular Misiones Portuguese exhibits high—nearly categorical rates of depleted plural marking, including invariant consonant-final nouns and adjectives. In the author's corpus of vernacular Misiones Portuguese, representing forty-eight Misiones natives the average rate of depleted plural marking was 94.6 %, with some environments (e.g., after a numerical quantifier) yielding depletion rates as high as 100 % (Lipski 2017). Quantitatively this places Misiones Portuguese at the extreme lower end of the continuum of variation of Brazilian Portuguese plural marking.

(8) Invariable plural consonant-final nouns and adjectives (also found in northern Uruguay):

os *ladrão* viste de fantasma 'the thieves are disguised as ghosts' (San Antonio)
não sei que outros *canal*; tem lindos *canal* argentino 'I don't know what other channels, there are nice Argentine channels' (El Soberbio)
da dez mil peso pra as *mulher* embarazada 'they give 10,000 pesos to pregnant women' (Comandante Andresito)

(9) Use of the third person singular verb instead of first person plural and third person plural: *nós mora[mos]* 'we live,' *êles fala[m]* 'they live.' This feature is found in northern Uruguay and vernacular Brazilian Portuguese (e.g., Azevedo 1989, 865; Guy 1981, 234s.; Lucchesi 1998, 85s.). The replacement of first-person plural verb forms by invariant third-person singular forms predominates in the imperfect tense (e.g. Naro et al. 1999, 203; Rodrigues 2004, 123–124; Seara 2000, 183–184; Zilles et al. 2005, 204), possibly due to the tendency to avoid verbal forms with antepenultimate stress. In the Misiones vernacular corpus 99 % of first-person plural imperfective verbs were realized as third-person singular forms (*trabaiava*, *ia*, etc.), reflecting the nearly categorical status of these non-canonical variants in Misiones Portuguese as opposed to the greater variability found in Brazil.

(10) In the vernacular Portuguese of Misiones the third person singular verb form is sometimes used for first-person singular reference. In Brazil this is not found, except in the semi-creole Afro-Brazilian dialect of Helvécia (Lucchesi/Baxter/Ribeiro 2009):

eu *mora* Iguazú também 'I live in Puerto Iguazú too' (Bernardo de Irigoyen)
agora não *sabe* eu dizer 'now I don't know what to say' (Irigoyen)
eu *tein* un cavalinho 'I have a pony' (Puerto Iguazú)
eu *tein* dos irmão (El Soberbio)
eu *tein* doze irmão 'I have twelve siblings' (Alba Posse)
algo *tein* [tenho...jeito para as matemáticas] 'I have some [talent for mathematics]' (San Javier)

(11) In first-conjugation verbs (*-ar*) the first-person plural forms frequently end in *-emo* instead of canonical Portuguese *-amos*; this feature is also characteristic of northern Uruguay and the vernacular Portuguese of the neighboring Brazilian states.

trabalhemo na chacra 'We work on the farm' (Bernardo de Irigoyen)
nos *moremo* perto do povo 'we live near town' (San Antonio)
*brinquemo*em casa 'we play at home' (Puerto Iguazú)
*plantemo*fumo 'We plant tobacco' (El Soberbio)
não onde nós *moremo* mais em outra chacra 'Not where we live but on other farms' (Comandante Andresito)

In Misiones Portuguese 100 % of first-person plural verb forms of verbs in *-ar* were realized as *-emo*, rates much higher than the 50 %–60 % rates found in southern Brazil (Zilles et al. 2005). In first-conjugation verbs the theme vowel /a/ is also frequently replaced by [e] in third-person plural forms:

meus irmão *cuidem* a chacra 'my sibilings take care of the farm' (El Soberbio)
êles *levem* o produto pra lá 'they take the product there' (Comandante Andresito)

11 Brazilian Portuguese in northern Uruguay

The most widely studied trans-border Brazilian Portuguese-speaking area is northern Uruguay, where a variety known to linguists as *Fronterizo* 'border speech' or *dialectos portugueses del Uruguay* 'Uruguayan dialects of Portuguese' and by the speakers themselves as *Portuñol* is widely used. Dialects of Uruguayan Portuguese or *Fronterizo* are spoken all along the Uruguay-Brazil border and in some towns well into the interior of Uruguay, but two Uruguayan border cities have been the principal focus of sociolinguistic research. The first is Rivera (approximate population 110,000), and its sister city Santana do Livramento (pop. about 83,000). The two cities form a single urban mass; the international border winds its way through the middle of the two cities, but is not marked except for a few monuments. There are no border controls or other visible presence of an international border (except for changes in street signs) and residents freely travel between the two countries. Rivera has a duty-free shopping zone that attracts many Brazilian tourists, and Portuguese is heard throughout Rivera, both as spoken by Brazilians and by native Uruguayans. Uruguayans listen to Brazilian radio and television and many have attended Brazilian schools. More recently, Portuguese language classes have been implemented in Rivera. On the Brazilian side no Spanish is spoken, although it is certainly understood. The local variety of Portuguese spoken in Santana do Livramento is in general closer to urban Brazilian standards than the rustic speech that forms the basis for Uruguayan *Fronterizo,* although some non-standard sociolects can still be found on the Brazilian side of the border (e.g., Meirelles 2007; also Hensey 1966). Some 180 km. to the west of Rivera is the Uruguayan city of Artigas (pop. about 42,000), which is separated from its Brazilian counterpart Quaraí (pop. around 25,000) by the Cuareim River. A half-mile long bridge joins the two cities, and vehicular and pedestrian traffic passes freely and without border controls. As with the case of Rivera, the relatively greater economic development on the Brazilian side traditionally resulted in Uruguayans' closer ties to Brazil—including the use of Portuguese—than to the central Uruguayan government in Montevideo. Working class residents of Artigas and the surrounding towns have traditionally preferred to speak Portuguese (or *Portuñol*), although the community is increasingly shifting to predominantly Spanish-based language (Douglas 2004).

In one of the first linguistic studies, Rona (1965, 12) felt that *Fronterizo* was not a stable language, but rather a dynamic situation in which Spanish and Portuguese freely combined:

> 'A process of selection can be observed, but only in incipient idiolectal fashion, according to the individual preferences of each speaker. Naturally this does not mean that the selection always takes place between the Spanish **system** and the Portuguese **system**. It's not a matter of choosing between Spanish and Portuguese, but rather a double set of possibilities that are simultaneously available to each speaker, and between which the speaker can choose in discourse one or the other.'[2]

He insists (Rona 1965, 13) that the selection process is both voluntary and conscious. Later researchers such as Elizaincín (1973; 1976; 1979; 1992), Hensey (1972; 1982a; 1982b), Elizaincín/Behares (1981), Elizaincín/Behares/Barrios (1987), Carvalho (2003a; 2003b; 2004a; 2004b; 2006a; 2006b), Douglas (2004), and Waltermire (2006) regard *Fronterizo* speech as a language in its own right, subject to regional and social variation, tracing of isoglosses, and diachronic change.

Until 1862 the northern region of what is now Uruguay was a disputed territory populated almost entirely by Brazilian squatters. Beginning in 1862 the Uruguayan government began a deliberate settlement effort, sending internal colonists from the populated south in order to establish de facto occupancy of the northern border. According to all available historical records, only Portuguese was spoken in this region until well into the second half of the 19th century (Elizaincín 1992, 99–100). In 1877 the *Ley de Educación Común* 'common education law' was passed; among its provisions was the stipulation that all educational instruction was to take place in the "national language," i.e., Spanish. In the capital and the interior of the country, this law had the effect of displacing immigrant languages after the first generation, and creating an essentially monolingual society. Along the northern border, on the other hand, the failure to acknowledge Portuguese in any sphere of public or private life and the exclusive promotion of Spanish led to a situation best described as diglossic, with Uruguayan Portuguese/*Fronterizo* dialects becoming the "low" variety, deprived of any possibility of standardization or expansion. Behares (1985) characterizes the survival of the non-prestigious hybrid *Fronterizo* dialects in northern Uruguay as the unexpected and presumably undesired results of poorly thought-out language planning (see also Carvalho 2006b).

---

2 "Puede, sí, apreciarse un proceso de selección, pero apenas en un grado incipiente, es decir, casi individual, idiolectal, según las preferencias de cada individuo hablante. Esto no significa, naturalmente, que la selección ocurra siempre entre el **sistema** portugués y el **sistema** castellano. No se trata de elegir entre hablar en portugués o hablar en castellano. Se trata más bien de un doble juego de posibilidades que están simultáneamente a disposición de cada hablante, y entre las cuales puede elegir, en el discurso, ya unas ya otras."

There is little accurate information on the nature of *Fronterizo* before the late 1960's, but judging by indirect references to language usage in northern Uruguay, including literary references (it appears that *Fronterizo* had stabilized by the first quarter of the 20th century and remained relatively stable for several generations thereafter. Since the late 1980's, there is evidence of an increasing tendency for natives of northern Uruguay to use more or less standard Spanish with one another, with spontaneous *Fronterizo* usage gradually becoming identified with the lower working classes and rural communities (Azevedo 2000; Giuffra 1900; Lipski 2008). As for ongoing sources of lexicalization, contemporary *Fronterizo* draws on both Spanish (e.g., from the schools, the local university, and the increasing presence of Spanish-language media), and Portuguese (principally from Brazilian television).

Most individuals who fluently speak *Portuñol* can also switch entirely to the regional variety of Spanish, but not to other varieties of Portuguese: the bilingualism is not Spanish-Portuguese, but rather Spanish-*Portuñol*, given that the latter language is grammatically grounded in Portuguese. Northern Uruguayan Portuguese shares many of the same traits as in Misiones, Argentina, although as noted by Carvalho (2004a) there is increasing sociolinguistic pressure to emulate educated Brazilian Portuguese, or to switch entirely to Spanish. Partially offsetting these trends there is an emerging published literary and culinary tradition in *Fronterizo* (e.g., Behares/Díaz 1998; Behares/Díaz/Holzmann 2004), as well as many albums of popular music by the Rivera artist Chito de Mello, with many songs in "Portuñol," some of which even purport to explain its usage.

Among the most extensively documented features of vernacular Uruguayan Portuguese are the following:

(1) Both Spanish and Portuguese articles appear in *Fronterizo*, a fact that is facilitated by the minimal differences between Spanish *los*, *la* and *las* and Portuguese *os*, *a* and *as*, respectively. Sometimes this results in combining a Spanish word with a Portuguese article or vice versa; on other occasions, both a Spanish and a Portuguese article may appear in a single sentence (Elizaincín et al. 1987, 41):

u [= *o*] material que se utiliza en *el* taier 'the materials that are used in the shop'
la importasão de automóviles 'the importation of autos'

(2) "Stripped" plural noun phrases (plural /-s/ marked only on the first element) are common in vernacular Uruguayan Portuguese. This trait is nearly categorical frequent in *Fronterizo*, even when Spanish articles are involved, and can even be found in vernacular Spanish of the border region (Carvalho 2006a; Lipski 2006). Some examples are (Elizaincín et al. 1987, 41s.):

Tein *umas vaca* para tirá leite 'I have some cows for milk'
Saí cum trinta y sei *gol* 'I scored 36 goals'

(3) Among *Fronterizo* speakers, combinations like *nós tinha* 'we had' [standard Ptg. *nós tínhamos*] instead of *nosotros teníamos* may be heard (Rona 1965, 12; Elizaincín et al. 1987). Significantly, there are no instances of this gravitation towards the 3rd person singular as quasi-invariable verb stem in *Fronterizo* verbs of Spanish origin. *Fronterizo* speakers occasionally employ the third person singular in Portuguese verbs, instead of the first person singular, as also occasionally found in Misiones, Argentina:

entonci yo no tein [tenho] ese dinheiro 'then I don't have that money.'
Uruguayan Portuguese speakers also typically employ the suffix *-emo* for first-person plural first-conjugation (*-ar*) verbs, as in Misiones and in rural areas of Rio Grande do Sul state.

(4) As in Misiones and rural Rio Grande do Sul state, Uruguayan Portuguese did not palatalize /t/ and /d/ before /i/, but Carvalho (2004a) has documented an increasing tendency for Uruguayan speakers to emulate this trait as heard on Brazilian radio and television, and representing most urban areas in Brazil.

12 Brazilians in the diaspora

Although Brazil continues to receive immigration from other nations, in recent decades there has been a significant outward migration of Brazilians, especially to the United States. The greatest concentration of Brazilians in the United States is found in south Florida, where as many as 300,000 Brazilians and their descendants reside. This is the only region of the United States in which the Brazilian population is cohesive enough to constitute a mini-speech community. This relatively recent immigration has yet to result in a cohesive Brazilian-American variety of Portuguese, but there is now generations of heritage Portuguese speakers whose language reflects some influence of English, for example preposition stranding (Lemos 2013).

Beginning in the 1980's many Brazilians of Japanese descent emigrated to Japan, where they worked as contract laborers (*Dekasegi*) and continue to speak Portuguese as well as Japanese. Today there are some 300,000 Japanese-Brazilians in Japan (the second-largest Brazilian diasporic community), many of whom engage in code-switching emblematic of their hybrid identity (e.g. Nakamizu 2003). Similar code-switching is also found among some Japanese-Brazilians in Brazil (Nawa 1988).

The third-largest Brazilian diasporic community is found in neighboring Paraguay (Albuquerque 2009; Blanc 2015; Zaar 2001). Given frequent travel between the two nations, most Brazilians in Paraguay continue to speak Portuguese, some Spanish, and almost no Guaraní. The *Brasiguaios*, itinerant Brazilians who frequently move between the two countries, do show some admixture with Spanish and occasionally

Guaraní when speaking Portuguese (Santos 2004; Santos/Cavalcanti 2008; Dalinghaus 2009; Spirandel 2006; Thun 2004).

13 Conclusion

Brazilian Portuguese has been enriched by the variegated tapestry of languages resulting from immigration, and which have substantially nuanced the speech of numerous rural communities. At the micro-dialectal level many traces of these languages can still be detected in the Portuguese of bilingual, heritage, and even monolingual speakers. As immigration continues, and as the Brazilian diaspora expands its reach, additional contact-induced micro-dialectal variation in Brazilian Portuguese is to be expected.

14 References

Adachi, Nobuko (1997), *Language choice: an ethnographic study of the Nôhon-Shugi ideology and language maintenance of Japanese immigrants in Brazil*, Doctoral thesis, Toronto (ON), University of Toronto.

Adachi, Nobuko (2001), *Japanese Brazilians: the Japanese language communities in Brazil,* Studies in the Linguistic Sciences 31/1, 161–178.

Albuquerque, José Lindomar C. (2009), *A dinâmica das fronteiras: deslocamento e circulação dos "brasiguaios" entre os limites nacionais*, Horizontes Antropológicos 15/31, 137–166.

Alessi, Mariana Longhi Batista (2013), *A migração de Haitianos para o Brasil*, Conjuntura Global 2/2, 82–86.

Almeida, Tainá Aragão de/Santi, Vilso Junior (2018), *Somos migrantes: o uso das redes sociais na produção midiática alternativa sobre a migração venezuelana em Roraima*, Aturá; Revista Pan-Amazônica de Comunicação 2/1, 136–156.

Altenhofen, Cléo Vilson (1996), *Hunsrückisch in Rio Grande do Sul: ein Beitrag zur Beschreibung einer deutschbrasilianischen Dialektvarietät im Kontakt mit dem Portugiesischen*, Stuttgart, Steiner.

Altenhofen, Cléo V./Frey, Jaqueline (2006), *Das bresilionische Deitsch unn die deitsche Bresilioner: en Hunsrickisch Red fo die Sprocherechte*, Revista Contingentia 1/1, 39–50.

Alves, Ieda Maria/Maroneze, Bruno Oliveira (2004), *Italianismos na língua portuguesa contemporânea*, Revista de Italianística 9, 29–36.

Andrade, Julieta de (1984), *Cultura crioula e Lanc-Patuá no norte do Brasil*, São Paulo, Escola de Folclore.

Armilato, Tales Giovani (2012), *A comunicação no rádio e a preservação de uma identidade lingüística regional: o talian*, Master's thesis, Caxias do Sul Universidade de Caixas do Sul.

Azevedo, Milton (1989), *Vernacular features in educated speech in Brazilian Portuguese*, Hispania 72, 862–872.

Azevedo, Milton (2000), *"Un lenguaje pintoresco y especial": an early instance of Portuguese-Spanish speech,* Revista Portuguesa de Humanidades 4, 95–109.

Bandeira, Marta Tessmann/Zimmer, Márcia C. (2012), *The dynamics of interlinguistic transfer of VOT patterns in multilingual children*, Linguagem & Ensino 15/2, 341–364.

Barros, Fernando Hélio Tavares de/ Philippsen, Neusa Inês (2013), *O Hunsrückisch: entre o Hochdeutsch e o português brasileiro: o caso do rádio na comunidade teuto-gaúcha norte mato-grossense*, Revista de Letras Norte@mentos 6/12, 111–132.

Behares, Luis (1985), *Planificación lingüística y educación en la frontera uruguaya con Brasil*, Montevideo, Instituto Interamericano del Niño.

Behares, Luis/Díaz, Carlos (edd.) (1998), *Os som de nossa terra: productos artístico-verbales fronterizos,* Montevideo, Universidad de la República/UNESCO.

Behares, Luis/Díaz, Carlos/Holzmann, Gerardo (2004), *Na frontera nós fizemo assim: lengua y cocina en el Uruguay fronterizo*, Montevideo, Universidad de la República.

Beilke, Neubiana Silva Veloso (2013a), *Pomerano: uma variedade germânica em Minas Gerais*, Anais do SILEL 3/1, 1–14.

Beilke, Neubiana Silva Veloso (2013b), *Do nativo ao pomerano: as línguas, os dialetos e falares vivos de um Brasil pouco conhecido*, Domínios de Lingu@gem 7/1, 263–283.

Benincá, Ludimilla Rupf (2009), *Dificuldade no domínio de fonemas do português por crianças bilíngues de português e pomerano*, Revista (Con) textos Linguísticos 3/3.1, 53–70.

Blanc, Jacob (2015), *Enclaves of inequality: Brasiguaios and the transformation of the Brazil-Paraguay borderlands*, Journal of Peasant Studies 42/1, 145–158.

Blank, Marceli Tessmer (2013), *Aspectos fonético-fonológicos da adquisição da escrita do português por crianças bilingües (pomerano/português)*, Master's thesis, Pelotas, Universidade Federal de Pelotas.

Borges, Maria Isabel (2007), *Um caso de variação de gênero em Santa Catarina: contato lingüístico entre o português e o italiano*, Revista Virtual de Estudos da Linguagem 5/9, 1–26.

Boruszenko, Oksana (1981), *A imigração ucraniana*, München, Klett-Cotta.

Boso, Ivette Marli (1991), *Entre passado e futuro: bilingüismo em uma comunidade trentino-brasileiro*, Master's thesis, Florianópolis, Universidad Federal de Santa Catarina.

Bovo, Nínive M. P. (2004), *A variação da vibrante e seu valor social*, Master's thesis, Caxias do Sul, Universidad de Caxias do Sul.

Bremenkamp, Elizana Schaffel (2014), *Análise sociolingüística da manutenção da língua pomerana em Santa Maria de Jetibá, Espírito Santo*, Master's thesis, Vitória, Universidade Federal do Espírito Santo.

Brescancini, Cláudia/Monaretto, Valéria Neto de Oliveira (2008), *Os róticos no sul do Brasil: panorama e generalizações*, Signum: Estudos Linguisticos 11/2, 51–66.

Bullock, Barbara/Toribio, Almeida Jacqueline (edd.) (2009), *The Cambridge handbook of linguistic code-switching*, Cambridge, Cambridge University Press.

Campos, Cynthia Machado (2006), *A política da língua na era Vargas: proibição do falar alemão e resistências no Sul do Brasil*, Campinas, Editora da UNICAMP.

Caprara, Loredana de Stauber/Mordente, Olga Alejandra (2004), *Panorama dell' italiano in San Paolo nel contesto plurilinguistico brasiliano*, Revista de Italianística 9, 101–112.

Carvalho, Ana Maria (2003a), *Rumo a uma definição do Português Uruguaio*, Revista Internacional de Lingüística Iberoamericana 1/2, 125–149.

Carvalho, Ana Maria (2003b), *The sociolinguistic distribution of (lh) in Uruguayan Portuguese: a case of dialect diffusion*, in: Silvina Montrul/Francisco Ordóñez (edd.), *Linguistic theory and language development in Hispanic languages*, Somereville (MA), Cascadilla Press, 30–44

Carvalho, Ana Maria (2004a), *I speak like the guys on TV: palatalization and the urbanization of Uruguayan Portuguese*, Language Variation and Change 16, 127–151.

Carvalho, Ana Maria (2004b), *Diagnóstico sociolingüístico de comunidades escolares fronterizas en el norte de Uruguay*, in: Claudia Brovetto/Javier Geymonat (edd.), *Portugués del Uruguay y educación bilingüe*, Montevideo, Administración Nacional de Educación Pública, 44–96.

Carvalho, Ana Maria (2006a), *Nominal number marking in a variety of Spanish in contact with Portuguese*, in: Timothy L. Face/Carol A. Klee (edd.), *Selected Proceedings of the 8th Hispanic Linguistics Symposium*, Somerville, MA, Cascadilla Proceedings Project, 154–166.

Carvalho, Ana Maria (2006b), *Políticas lingüísticas de séculos passados nos dias de hoje: o dilema sobre a educação bilíngüe no norte do Uruguai*, Language Problems and Language Planning 30, 149–171.

Castro, Giliola Maggio de (1997), *Pedrinhas Paulista: comunidade italiana que ainda permanece ligada à língua materna*, Revista de Italianística 5/5, 275–280.

Cenni, Franco (2003), *Italianos no Brasil: "andiamo in 'Merica…"*, São Paulo, Editorial da Universidade de São Paulo.

Choi, Keum Joa (1991), *Além do arco-íris: a imigração coreana no Brasil*, Master's thesis, São Paulo, University of São Paulo.

Choi, Keum Joa (1998), *Imigração coreana na cidade de São Paulo*, Revista do Instituto de Estudos Brasileiros 40, 233–238.

Cogo, Denise (2014), *Haitianos no Brasil. Comunicação e interação em redes migratórias transnacionais*, Chasqui: Revista Latinoamericana de Comunicación 125, 23–32.

Costa, Luciane Trennephol da/Gielinski, Márcia Inês (2014), *Detalhes fonéticos do polonês falado em Malllet*, Revista (Con) textos Linguísticos 8/10, 159–174.

Cotinguiba, Marília Lima Pimentel/Cotinguiba, Geraldo Castro (2014a), *Wout, raketè, fwontyè, anpil mizè: reflexões sobre os limites da alteridade em relação à imigração haitiana para o Brasil*, Universitas: Relações Internacionais 12/1, 73–86.

Cotinguiba, Marília Lima Pimentel/Cotinguiba, Geraldo Castro (2014b), *Imigração haitiana para o Brasil: os desafios no caminho da educação escolar*, Pedagógica: Revista do programa de Pós-graduaçao em Educaçao-PPGE 16/33, 61–88.

Couto, Hildo do (2011), *Contato entre português e espanhol*, in: Heliana Mello/Cléo Altenhofen/Tommaso Raso (edd.), *Os contatos linguísticos no Brasil*, Belo Horizonte, Editora UFMG, 360–395.

Cymbalista, Renato (2007), *A comunidade boliviana em São Paulo: definindo padrões de territorialidade*, Cadernos Metrópole 17, 119–133.

Dalinghaus, Ione Vier (2009), *Alunos brasiguaios em escola de fronteira Brasil/Paraguai: um estudo linguístico sobre aprendizagem do português em Ponta Porã, MS*, Master's thesis, Paraná, Universidade Estadual do Oeste do Paraná.

Dal Picol, Greyce (2013), *A morfossintaxe na oralidade do vêneto sul-rio-grandense: perfil dialetal de comunidades rurais da região da 4ª légua*, Master's thesis, Caxias do Sul, Universidade de Caxias do Sul.

Damke, Ciro (1997), *Sprachgebrauch und Sprachkontakt in der deutschen Sprachinsel in Südbrasilien*, Frankfurt am Main, Lang.

Damke, Ciro (2008), *O Brasildeutsch como fator de conservação da língua alemã no Brasil*, Revista Trama 4/7, 115–123.

Daviña, Liliana Silvia (2003), *Fronteras discursivas en una región plurilingüe: español y portugués en Misiones*, Master's thesis, Buenos Aires, Universidad de Buenos Aires.

Day, Kelly Cristina Nascimento (2013), *Fronteiras linguísticas e fronteiras políticas: relações linguísticas e socio-históricas na fronteira do Brasil com a Guiana Francesa*, Cadernos de Letras da UFF 23/47, 163–182.

Day, Kelly Cristina Nascimento (2017), *A ilha Amapá-panorama ecolinguístico da fronteira franco-brasileira*, Cadernos de Letras da UFF 26/53, 99–120.

Douglas, Kendra (2004), *Uruguayan Portuguese in Artigas: tri-dimensionality of transitional local varieties in contact with Spanish and Portuguese standards*, Doctoral thesis, Madison (WI), University of Wisconsin.

Druszcz, Arlindo Milton (1984), *O bilingüismo em Araucária: A interferência polonesa na fonologia portuguesa*, Master's thesis, Paraná, Universidade Católica do Paraná.
Dück, Elvine Siemens (2011), *Vitalidad lingüística do Plautdietsch em contato com variedades standard faladas em comunidades menonitas no Brasil*, Doctoral thesis, Porto Alegre, Universidade Federal de Rio Grande do Sul.
Dück, Elvine Siemens (2012), *Vitalidad lingüística do Plautdietsch no Brasil*, in: *Anais do X Encontro do CELSUL, Cascavel-PR, 24 a 26 de outubro de 2012*, https://docplayer.com.br/12662576-Vitalidade-linguistica-do-plautdietsch-no-brasil.html, last accessed 18.02.2022.
Dvorak, Anna (2013), *A hidden immigration: the geography of Polish-Brazilian cultural identity*, Doctoral thesis, Los Angeles (CA), University of California Los Angeles.
Elizaincín, Adolfo (1973), *Algunos aspectos de la sociolingüística del dialecto Fronterizo*, Montevideo, Universidad de la República.
Elizaincín, Adolfo (1976), *The emergence of bilingual dialects on the Brazilian-Uruguayan border*, International Journal of the Sociology of Language 9, 123–134.
Elizaincín, Adolfo (1979), *Algunas precisiones sobre los dialectos portugueses en el Uruguay*, Montevideo, Universidad de la República.
Elizaincín, Adolfo (1992), *Dialectos en contacto: español y portugués en España y América*, Montevideo, Arca.
Elizaincín, Adolfo/Behares, Luis (1981), *Variabilidad morfosintáctica de los dialectos portugueses del Uruguay,* Boletín de Filología 31, 401–417.
Elizaincín, Adolfo/Behares, Luis/Barrios, Graciela (1987), *Nos falemo brasilero*, Montevideo, Editorial Amesur.
Espiga, Jorge Walter da Rocha (1997), *Influência do espanhol na variação da lateral pós-vocálica do português da fronteira*, Master's thesis, Pelotas, Universidade Católica de Pelotas.
Espiga, Jorge Walter da Rocha (2002), *A lateral posvocálica na fronteira dos Campos Neurais: estudo sociolinguístico da regra telescópica nos dialetos de Chuí e Santa Vitória do Palmar*, Letras de Hoje 37/1, 49–68.
Espiga, Jorge Walter da Rocha (2003), *Alofonia de /l/ no sul do Rio Grande do Sul: aspectos fonéticos e fonológicos*, in: Demerval da Hora/Gisela Collischonn (edd.), *Teoria lingüística: fonologia e outros temas*, João Pessoa, Universitária/Universidade Federal da Paraíba, 251–290.
Espiga, Jorge Walter da Rocha (2006), *O contato do português com o espanhol no sul do Brasil*, in: Paulino Vandresen (ed.), *Variação, mudança e contato lingüístico no português da região sul,* Pelotas, EDUCAT, 261–279.
Faggion, Carmen Maria (2006), *Harmonia vocálica com -inho e -zinho: uma marca dialetal específica*, Revista Virtual de Estudos da Linguagem—ReVEL 7, 1–19.
Faggion, Carmen Maria (2010), *O português popular falado por ítalo-descendentes do sul do Brasil,* in: Maria João Marçalo et al. (edd.), *Língua portuguesa: ultrapassar fronteiras, juntar culturas*, Évora, Universidade de Évora, 44–81.
Ferraz, Aderlande Pereira (2007), *O panorama lingüístico brasileiro: a coexistência de línguas minoritárias com o português*, Filologia e Linguística Portuguesa 9, 43–73.
Fraga, Letícia (2009a), *O bilinguismo português/holandês em Carambeí/PR*, Todas as Letras 11/1, 82–91.
Fraga, Letícia (2009b), *A variedad de português falada pelos "holandeses" de Carambeí: identidade e r-forte*, Revista Brasileira de Linguística 9, 621–645.
Fraga, Letícia (2011), *Os "holandeses" de Carambeí e sus línguas: um estudo sobre identidade*, Estudos Linguísticos 40, 1084–1098.
Fritzen, Maristela Pereira (2007), *"Ich kann mein Name mit letra junta und letra solta schreib": deutsch-portugiesisches Code-Switching in einer Grundschule im südbrasilianischen Immigrationsgebiet*, Pandaemonium Germanicum 11, 125–156.

Fritzen, Maristela Pereira (2008), *Línguas em contato/conflito: em foco uma escola rural em zona de imigração alemã no sul do Brasil*, Anais do CELSUL 8, https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwjVh5H57ZHwAhWLTN8KHV4OBlgQFjAAegQIAxAD&url=http%3A%2F%2Fwww.leffa.pro.br%2Ftela4%2FTextos%2FTextos%2FAnais%2FCELSUL_VIII%2Flinguas_em_contato_conflito.pdf&usg=AOvVaw23C0DT5vGCPhO0feaB1eC4, last accessed 18.02.2022.

Frosi, Vitalina Maria (1987), *I dialetti italiani nel Rio Grande do Sul e il loro sviluppo nel contesto socioculturale ed economico: prevalenza del dialetto veneto*, in: Vicenzo Locaschio (ed.), *L'italiano in America Latina*, Firenze, Le Monnier, 136–163.

Frosi, Vitalina Maria/Faggion, Carmen Maria/Dal Corno, Giselle Olivia Mantovani (2005), *Da estimatização à solidaridade: atitudes lingüísticas na RCI*, MÉTIS: História e Cultura 4/8, 257–280.

Frosi, Vitalina Maria/Faggion, Carmen Maria/Dal Corno, Giselle Olivia Mantovani (2007), *Linguagem da região de colonização italiana do nordeste do Rio Grande do Sul: prestígio e estigmatização*, Revista Virtual de Estudos da Linguagem-ReVEL 5/9.

Frosi, Vitalina Maria/Faggion, Carmen Maria/Dal Corno, Giselle Olivia Mantovani (2008), *Prestígio e estigmatização: dialeto italiano e língua portuguesa da região de colonização italiana do nordeste do Rio Grande do Sul*, Revista da ABRALIN 7/2, 139–167.

Frosi, Vitalina Maria/Mioranza, Ciro (1983), *Dialetos italianos*, Caxias do Sul, EDUCS.

Fuchs, Cristina Yukie Miyaki (1996), *Interferências fonológicas nos falantes bilingües do português e do japonês: fatores socio e psicolingüísticos*, Master's thesis, Curitiba, Universidade Federal do Paraná.

Fujiwara, Érika Yurie (2014), *A criação de neologismos de base japonesa por falantes de português*, Master's thesis, Três Lagoas, Universidade Federal de Mato Grosso do Sul.

Garcia, Neiva Mara Zanin (2009), *Estudo linguístico-etnográfico em comunidade paranaense de imigrantes ucranianos: do passado ao presente*, Signum: Estudos da Linguagem 12/1, 163–189.

Gibo, Lucila Etsuko (2014), *O contato lingüístico na comunidade okinawana do Brasil e o português falado pelos okinawanos*, Encontros Lusófonos 16, 5–20.

Giuffra, Santiago (1900), *Fronterizas: paliques uruguayos-brasileros*, Montevideo, Imprenta El Siglo Ilustrado.

Godoy, Gabriel (2011), *O caso dos haitianos no Brasil e a via da proteção humanitária complementar*, in: André de Carvalho Ramos/Gilberto Rodrigues/Guilherme Assis de Almeida (edd.), *60 anos de ACNUR*, São Paulo, Editora CLA Cultural, 45–68.

Guy, Gregory (1981), *Linguistic variation in Brazilian Portuguese: aspects of the phonology, syntax, and language history*, Doctoral thesis, Philadelphia (PA), University of Pennsylvania.

Hensey, Fritz (1966), *Livramento/Rivera: the linguistic side of international relations*, Journal of Inter-American Studies 8/4, 520–534.

Hensey, Fritz (1972), *The sociolinguistics of the Brazilian-Portuguese border*, The Hague, Mouton.

Hensey, Fritz (1982a), *Uruguayan fronterizo: a linguistic sampler*, Word 33, 193–198.

Hensey, Fritz (1982b), *Spanish, Portuguese and Fronteriço: languages in contact in northern Uruguay*, International Journal of the Sociology of Language 34, 9–23.

Höhmann, Beate/Savedra, Mônica Maria Guimarães (2011), *Das Pommerische in Espírito Santo: Ergebnisse und Perspektiven einer soziolinguistischen Studie*, Pandaemonium Germanicum 18, 283–300

Jye, Chen Tsung/Shyu, David Jye Yuan/ Menezes Jr, Antonio José Bezerra de (2006), *Os imigrantes chineses no Brasil e a sua língua*, Synergies Brésil 7, 57–64.

Kanazawa, Hiroki/Loveday, Leo (1988), *The Japanese immigrant community in Brazil: language contact and shift*, Journal of Multilingual and Multicultural Development 9, 423–435.

Kaufmann, Göz (2009), *Falar espanhol or hablar português: attitudes and linguistic behavior on the Brazilian-Uruguayan and Brazilian-Argentinian borders*, Romanistisches Jahrbuch 60, 276–317.
Kim, Han Chul (2005), *Aquisição do artigo definido em português como segunda língua por aprendizes coreanos*, Doctoral thesis, Porto Alegre, Universidade Federal do Rio Grande do Sul.
Krug, Marcelo Jacó (2004), *Identidade e comportamento lingüístico na percepção da comunidade plurilíngüe alemão-italiano-português de imigrante—RS*, Master's thesis, Porto Alegre, Universidade Federal do Rio Grande do Sul.
Lara, Claudia Camila/Battisti, Elisa (2014), *O Voice Onset Time das plosivas do português brasileiro em contato com o Hunsrückisch e seu desvozeamento variáve*l, Fórum Linguístico 11/1, 39–50.
Leão, Paula Biegelmeier (2007), *Transmissão intergeracional do alemão em contato com o português em Vale Real—RS*, Master's thesis, Porto Alegre, Universidade Federal do Rio Grande do Sul.
Lemke, Cibele Krause (2010), *A língua espanhola entre o português e o ucraniano: um estudo sobre a aprendizagem de terceiras línguas*, Revista Iberoamericana de Educación 52, 1–11.
Lemos, Simone H. de (2013), *Preposition stranding in heritage speakers of Brazilian Portuguese,* Master's thesis, University Park (FL), Florida International University.
Leopoldino, Everton Altmayer (2009), *A fala dos tiroleses de piracicaba: um perfil linguístico dos bairros Santana e Santa Olímpia*, Master's thesis, São Paulo, Universidade de São Paulo.
Leopoldino, Everton Altmayer (2014), *A fala dos tiroleses de piracicaba: um perfil linguístico dos bairros Santana e Santa Olímpia*, Doctoral thesis, São Paulo, Universidade de São Paulo.
Lipski, John (1985), *Linguistic aspects of Spanish-English language switching*, Tempe, Arizona State University, Center for Latin American Studies.
Lipski, John (2006), *Too close for comfort? The genesis of portuñol/portunhol*, in: Timothy L. Face/Carol A. Klee (edd.), *Selected proceedings of the 8th Hispanic Linguistics Symposium*, Somerville, MA, Cascadilla Press, 1–22, www.lingref.com, last accessed 18.02.2022.
Lipski, John (2008), *Searching for the origins of Uruguayan Fronterizo dialects: radical code-mixing as "fluent dysfluency",* Journal of Portuguese Linguistics 8, 5–46.
Lipski, John (2009), *"Fluent dysfluency" as congruent lexicalization: a special case of radical code-mixing,* Journal of Language Contact 2, 1–39.
Lipski, John (2010), *Spanish and Portuguese in contact*, in: Raymond Hickey (ed.), *Handbook of language contact*, New York, Wiley-Blackwell, 550–580.
Lipski, John (2011a), *Um caso de contato de fronteira: o sudoeste*, in: Heliana Mello/Cléo Altenhofen/Tommaso Raso (edd.), *Os contatos linguísticos no Brasil*, Belo Horizonte, Editora UFMG, 349–368.
Lipski, John (2011b), *Encontros lingüísticos fronteiriços*, Ideação 13/2, 83–100.
Lipski, John (2015), *Portuguese/portuñol in Misiones, Argentina: another Fronterizo?*, in: Sandro Sessarego/Melvin González (edd.), *New perspectives on Hispanic contact linguistics*, Frankfurt/Madrid, Vervuert/Iberoamericana, 233–263.
Lipski, John (2017), *Portuguese or Portuñol? Language contact in Misiones, Argentina*, Journal of Linguistic Geography 4, 47–64.
Lipski, John (2018a), *La interfaz portugués-castellano en Misiones, Argentina: zona de prueba para la alternancia de lenguas*, Estudios Filológicos 60, 169–190.
Lipski, John (2018b), *La evolución de la interfaz portugués-español en el noreste argentino*, in: Dolores Corbella/Alejando Fajardo (edd.), *Español y portugués en contacto: préstamos léxicos e interferencia*, Berlin/Boston, De Gruyter, 391–412.
Lipski, John (2019), *Field-testing code-switching constraints: a report on a strategic languages project*, Languages 4/1, 7, https://doi.org/10.3390/languages4010007.
Lipski, John (2020), *Portuguese-Spanish contacts in Misiones, Argentina: probing (for) code-switching constraints*, in: Luis Ortiz-López/Rosa Guzzardo Tamargo/Melvin González Rivera (edd.),

Hispanic contact linguistics: theoretical, methodological and empirical perspectives, Amsterdam/Philadelphia, Benjamins, 235–259.
Lobato, Lúcia Maria Pinheiro (2005), *Sobre a questão da influência ameríndia na formação do português do Brasil,* in: Denize Elena Garcia da Silva (ed.), *Língua, gramática e discurso*, Goiânia, Editorial/Grupo de Estudos da Linguagem do Centro-Oeste, 54–86.
López, Brenda V. de (1993), *Lenguaje fronterizo en obras de autores uruguayos*, Montevideo, Editorial Nordan-Comunidad.
Loriato, Sarah/Ponzo Peres, Edenize (2013), *Gênero e a realização do /r/ em uma situação de contato lingüístico*, Entretextos 13/2, 431–446.
Lownes, Steven Phillip (2018), *Johnny "Joãozinho" Reb: The Creation and Evolution of Confederate Identity in Brazil*, Doctoral thesis, Columbus (OH), Ohio State University.
Lucchesi, Dante (1998), *A constituição histórica do português brasileiro como um processo bipolarizador: tendências atuais de mudança nas normas culta e popular*, in: Sybille Grosse/Klaus Zimmermann (edd.), *"Substandard" e mudança no português do Brasil*, Frankfurt am Main, TFM, 73–99.
Lucchesi, Dante/Baxter, Alan/Ribeiro, Ilza (2009), *O português afro-brasileiro*. Salvador, EDUFBA.
Luersen, Rosane Werkhausen (2009), *A situação de contato plurilíngue no sul do Brasil*, Revista Científica da Faculdade Salesiana Maria Auxiliadora Visões 7, 70–87.
Machado, Ananda/Pereira, Jama Peres (2019), *Diagnóstico sociolinguístico nas escolas municipais após a sanção à Lei 211/2014, que cooficializou as línguas indígenas na sede do Município de Bonfim, Roraima*, Tellus 19/40, 137–161.
Maia, Ivene Carissini da (2004), *Intercambios lingüísticos de frontera: incidencia en el hablar de los alumnos del profesorado en portugués de la UNAM*, Master's thesis, Posadas, Universidad Nacional de Misiones.
Maia, Ivene Carissini da/Triches, Simone Maria (2017), *Dinâmica social da fronteira e seus efeitos na linguagem*, Web Revista SOCIODIALETO 7/20, 271–283.
Malczewski, Zdzisław (2010), *Polonii Brazylijskiej Obraz Wlasny,* Warsaw, Biblioteka Iberyjska.
Maltzahn, Paulo César (2013), *Construção e formação da identidade étnica do imigrante alemão no Rio Grande do Sul (BRASIL): A Língua Alemã como marcador importante de etnicidade*, http://www.abrapa.org.br/cd/pdfs/Maltzahn-S1-Paulo-korr.pdf, last accessed 18.02.2022.
Margotti, Felício Wessling (2004), *Difusão sócio-geográfica do português em comtato com o italiano no sul do Brasil*, Doctoral thesis, Porto Alegre, Universidade Federal do Rio Grande do Sul.
Medeiros, Regine del Negri (1982), *American Brazilian English*, American Speech 57, 150–152.
Meirelles, Virginia Andrea Garrido (2007), *Características fonético-fonológicas das variedades de espanhol e português faladas em Sant'ana do Livramento e Rivera*, Letras de Hoje 42/3, 169–177.
Meyer, Martina (2009), *Deitsch ou deutsch? macroanálise pluridimensional da variação do Hunsrückisch rio-grandense em contato com o português*, Undergraduate thesis, Porto Alegre, Universidade Federal de Rio Grando do Sul.
Mezavila, Albertina (2008), *Bilingüismo e as redes de comunicação na comunidade ucraniana de Cascavel*, Línguas e Letras 9/17, 289–304.
Miodunka, Władysław (2003), *Bilingwizm polsko-portugalski w Brazylii* [Polish-Portuguese bilingualism in Brazil.], Cracow, Universitas.
Montgomery, Michael/Melo, Cecil Ataide (1990), *The phonology of the lost cause: the English of the Confederados in Brazil*, English World-Wide 11/2, 195–216.
Montgomery, Michael/Melo, Cecil Ataide (1995), *The language: the preservation of southern speech among the colonists*, in: Cyrus B. Dawsey/James M. Dawsey (edd.), *The Confederados: Old South immigrants in Brazil*, Tuscaloosa, University of Alabama Press, 176–190.
Moraes, Isaias Albertin de (2013), *A imigração haitiana para o Brasil: causas e desafios*, Conjuntura Austral 4/20, 95–114.

Muysken, Pieter (2000), *Bilingual speech: a typology of code-mixing*, Cambridge, Cambridge University Press.
Nakamizu, Ellen (2003), *Language contact between Portuguese and Japanese: functions of code-switching in the speech of Brazilians living in Japan*, Bulletin of Portuguese-Japanese Studies 6, 73–91.
Naro, Anthony/Görski, Edair/Fernandes, Eulália (1999), *Change without change*, Language Variation and Change 11, 197–211.
Nawa, Takako (1988), *Bilingüismo e mudança de código. Uma proposta de análise com os nipo-brasileiros residentes em Brasília*, Master's thesis, Brasília, University of Brasília.
Niehoff, Stephanie (2011), *Bolivian immigrants in São Paulo: a sociolinguistic study of language contact in the city*, Doctoral thesis, Berlin, Freie Universität Berlin.
Nieto, Carlos (2014), *Migración haitiana a Brasil. Redes migratorias y espacio social transnacional*, Buenos Aires, CLACSO.
Ota, Junko (2009), *A língua falada nas comunidades rurais nipo-brasileiras do estado de São Paulo—considerações sobre koronia-go*, Synergies Brésil 7, 49–56.
Parlato-Oliveira, Erika, et al. (2010), *Plasticity of illusory vowel perception in Brazilian-Japanese bilinguals*, The Journal of the Acoustical Society of America 127/6, 3738–3748.
Peres, Edenize Ponzo (2011), *Aspectos da imigração italiana no Espírito Santo: a língua e cultura do Vêneto em Araguaia*, Dimensões 26, 44–59.
Pérez Gómez, Ricardo (2018), *¿Un caso de emigración cultural? Los sudistas confederados y el Brasil de su llegada (1860–1870)*, Procesos Históricos 34, 88–106.
Pertile, Marley Terezinha (2009), *O talian entre o italiano-padrão e o português brasileiro: manutenção e substituição no Alto Uruguai gaúcho*, Doctoral thesis, Porto Alegre, Universidade Federal do Rio Grande do Sul.
Piza, Douglas de Toledo (2012), *Um pouco da mundialização contada a partir da região da rua 25 de março: migrantes chineses e comércio "informal",* Master's thesis, São Paulo, Universidade de São Paulo.
Ponso, Letícia Cao (2003), *A variação do português em contato com o italiano na comunidade bilíngüe de São Marcos—RS*, Master's thesis, Porto Alegre, Universidade Federal de Rio Grande do Sul.
Prudente, Mabel Pettersen (2019), *As línguas wapichana, macuxi, português, inglês, creolese e espanhol nos cenários sociolinguísticos fronteiriços do Brasil com a República Cooperativa da Guiana*, Doctoral thesis, Goiânia, Universidade Federal de Goiás.
Putnam, Michael/Lipski, John (2016), *Null arguments in transitional trilingual grammars: field observations from Misionero German*, Multilingua 35, 85–104
Rezagli, Caroline Calderon (2009), *O português culto falado por alemães residentes na cidade de São Paulo: gramaticalização e contato entre línguas*, Master's thesis, São Paulo, Universidade de São Paulo.
Rickli, João Frederico (2003), *A "Comunidade da Bênção": religião, família e trabalho na Colônia Castrolanda,* Master's thesis, Paraná, Universidade Federal de Paraná.
Rodrigues, Angela C. S. (2004), *Concordância verbal, sociolingüística e história do português brasileiro*, Fórum Lingüístico (Florianópolis) 4/1, 111–145.
Rodrigues, Catarina Vaz (2009), *Bilinguismo no Espírito Santo: reflexos no português de adultos e crianças*, Signum: Estudos da Linguagem 12/1, 293–316.
Rona, José Pedro (1965), *El dialecto "Fronterizo" del norte del Uruguay*, Montevideo, Linardi.
Rosenberg, Peter (1998), *Deutsche Minderheiten in Lateinamerika*, in: Theo Haraden/Elke Hentschel (edd.), *Particulae particularum. Festschrift zum 60. Geburtstag von Harald Weydt*, Tübingen, Stauffenburg, 261–291.
Rosenberg, Peter (2002), *Deutsche Minderheiten in Lateinamerika*, Martius-Staden Jahrbuch 49, 9–50.

Rosenberg, Peter (2003a), *Vergleichende Sprachinselforschung: Sprachwandel in deutschen Sprachinseln in Russland und Brasilien*, Linguistik online 13, 273–323, https://doi.org/10.13092/lo.13.881.

Rosenberg, Peter (2003b), *Comparative speech island research: Some results from studies in Russia and Brazil*, in: William Keel/Klaus J. Mattheier (edd.), *Deutsche Sprachinseln weltweit: Interne und externe Perspektiven. German Language Varieties Worldwide: Internal and external Perspectives*, Frankfurt am Main, Lang, 199–238.

Rosenberg, Peter (2012), *Regularität und Irregularität in der Kasusmorphologie deutscher Sprachinselvarietäten (Russland, Brasilien)*, in: Klaus-Michael Köpcke/Andreas Bittner (edd.), *Regularität und Irregularität in Phonologie und Morphologie*, Berlin/Boston, Mouton de Gruyter, 177–218.

Rost, Cláudia Andrea (2008), *A identidade do teuto-brasileiro na região sul do Brasil*, Interdisciplinas 5/5, 215–234.

Saiki, Regina Akemi Uchima (2013), *Esboço sobre o ritmo do português-brasileiro / japonês: um estudo experimental*, Undergraduate thesis, Florianópolis, Universidade Federal de Santa Catarina.

Santos, Alessandra de Souza (2012), *Multilinguismo em Bonfim/RR: o ensino de Língua Portuguesa no contexto da diversidade linguística*, Doctoral thesis, Brasíslia, Universidade de Brasília.

Santos, Fabio (2017), *Re-mapping Europe. Field notes from the French-Brazilian borderland*, InterDisciplines. Journal of History and Sociology 8/2, 173–201.

Santos, Fernanda Naomi Zaphiro Pessoa/Vasconcelos, Thamires Marques (2016), *Venezuelanos no Brasil: da crise econômica para a crise política e midiática,* in: *Anais do XVII Encontro de História da Anpuh-Rio*, Rio de Janeiro, Anpuh-Rio, 1–16.

Santos, Maria Elena Pires (2004), *O cenario multilingue, multidialetal, multicultura de fronteira e o processo identitario brasiguaio na escola e no entorno social*, Campinas, Doctoral thesis, Campinas, Universidade Estadual de Campinas.

Santos, Maria Elena Pires/Cavalcanti, Marilda do Couto (2008), *Identidades híbridas, lingua (gens) provisórias–alunos brasiguaios em foco*, Trabalhos em Linguística Aplicada 47/2, 429–446.

Schaeffer, Shirlei Conceção Barth (2012), *Descrição fonetica e fonológica do pomerano falado no Espírito Santo*, Master's thesis, Vitória, Universidade Federal do Espírito Santo.

Schaumloeffel, Marco Aurélio (2003), *Estudo da interferência do português da variedade dialetal Hunsrück falada em Boa Vista do Herval*, Master's thesis, Paraná, Universidade Federal de Paraná.

Schissatti, Luciana Cristina Souza (2014), *Interrogativas—WH no dialeto trentino de Rodeio*, Doctoral thesis, Florianópolis, Universidade Federal de Santa Catarina.

Schneider, Maria Nilse (2012), *A manutenção do Hunsrückisch e do pomerano e o ensino de alemão em comunidades teuto-brasileiras no Rio Grande do Sul*, X Encontro do CELSUL–Círculo de Estudos Linguísticos do Sul, 1–15.

Seara, Izabel Christine (2000), *A variação do sujeto "nós" e "a gente" na fala florianopolitana*, Organon 14/28–29, 179–194.

Silva, Karen Kênnia Couto (2017), *A política linguística na região fronteiriça Brasil-Guiana Francesa: panorama e contradições*, Trabalhos Em Linguística Aplicada 56/2, 617–663.

Silva, Sidney Antonio da (2005), *A migração dos símbolos: diálogo intercultural e processos identitários entre os bolivianos em São Paulo*, São Paulo em perspectiva 19/3, 77–83.

Silva, Sidney Antonio da (2006), *Bolivianos em São Paulo: entre o sonho e a realidade*, Estudos avançados 20/57, 157–170.

Silva, Flávia Renata de (2010), *A educação alemã na colônia riograndense: 1922–1938 (Maracaí/Cruzália-SP)*, Master's thesis, São Paulo, Universidade Estadual Paulista UNESP de Presidente Prudente.

Simai, Szilvia/Baeninger, Rosana (2012), *Discurso, negação e preconceito: bolivianos em São Paulo*, in: Rosana Baeninger (ed.), *Imigração boliviana no Brasil*, Campinas, NEPO/UNICAMP, 195–210.

Simões, Gustavo da Frota (ed.) (2017), *Perfil Sociodemográfico e Laboral da Imigração Venezuelana no Brasil*, Curitiba, Editora CRV.

Souchaud, Sylvain (2010), *A imigração boliviana em São Paulo*, in: Ademir Pacelli Ferreira et al. (edd.), *Deslocamentos e reconstruções da experiência migrante*, Paris, Garamond, 267–290.

Souchaud, Sylvain/Baeninger, Rosana (2008), *Collas e cambas do outro lado da fronteira: aspectos da distribuição diferenciada da imigração boliviana em Corumbá, Mato Grosso do Sul*, Revista Brasileira de Estudos de População 25/2, 271–286.

Souza, Andréia Cristina (2012), *Crenças e atitudes lingüísticas: a língua alemã, a língua portuguesa e o Brasildeutsch*, in: *Anais 3º CIELLI: Colóquio de Estudos Lingüísticos e Literários, Universidade Estadual de Maringá, agosto 2014*, 220–231.

Spinassé, Karen Pupp (2008), *O hunsrückisch no Brasil: a língua como fator histórico da relação entre Brasil e Alemanha*, Espaço Plural 9/19, 117–126.

Spirandel, Marcia Anita (2006), *Brasileiros na fronteira com o Paraguai*, Estudos Avançados 20/57, 137–156.

Sturza, Eliana Rosa (1994), *O espanhol do cotidiano e o espanhol da escola: um estudo de caso na fronteira Brasil-Argentina*, Master's thesis, Santa Maria, Universidade Federal de Santa Maria.

Sufredini, Lourdes Claudete Schwade (1993), *Aspectos do bilingüismo alemão/português numa comunidade rural do oeste catarinense*, Master's thesis, Florianópolis, Universidade Federal de Santa Catarina.

Thun, Harald (2004), *O comportamento lingüístico dos brasiguaios no Paraguai visto a partir do material do Atlas Lingüístico Guarani-Románico (ALGR)*, in: Wolf Dietrich/Volker Noll (edd.), *O português do Brasil. Perspectivas da pesquisa atual*, Frankfurt/Madrid, Vervuert/Iberoamericana, 169–194.

Tobler, S. Joy (1983), *The grammar of karipúna creole*, Brasília, Summer Institute of Linguistics.

Tobler, Alfred W. (1987), *Dicionário crioulo karipúna/português português/karipúna*, Brasília, Summer Institute of Linguistics.

Tomiello, Marciana (2005), *A variação do ditongo nasal -ão como prática social no português de São Marcos/RS*, Master's thesis, Caxias do Sul, Universidad de Caxias do Sul.

Toscan, Mirian Peccat (2005), *O comportamento linguístico na comunidade bilingue ítalobrasileira de Nova Pádua/RS: identidade, prestígio e estigma linguísticos*, Master's thesis, Caxias do Sul, Universidade de Caxias do Sul.

Trento, Angelo (1989), *Do outro lado do Atlântico: um século de imigração italiana no Brasil*, São Paulo, Studio Nobel.

Valim, Hauley Silva/Berocan Veiga, Felipe/Vieira da Cunha, Neiva (2011), *O Saara Oriental: coreanos no Rio de Janeiro e as interfaces entre imigração, mercado e religião*, Estudos Afro-Asiáticos 33/1-2-3, 173–195.

Vandresen, Paulino (ed.) (2006), *Variação, mudança e contato lingüístico no português da região sul*, Pelotas, EDUCAT.

Vandresen, Paulino/Rodrigues Corrêa, Adriane (2008), *O bilingüismo Pomerano-Português na região de Pelotas, Português brasileiro II: contato linguístico, heterogeneidade e história*, Niterói, EdUFF, 39–51.

Vieira, Marcílio Melo (2010), *Para um estudo das influências fonológicas do italiano no português falado na cidade de São Paulo*, Master's thesis, São Paulo, Universidade de São Paulo.

Waltermire, Mark (2006), *Social and linguistic correlates of Spanish-Portuguese bilingualism on the Uruguayan-Brazilian border*, Doctoral thesis, Albuquerque (NM), University of New Mexico.

Weaver, Blanche Henry Clark (1961), *Confederate emigration to Brazil*, The Journal of Southern History 27, 33–53.

Zaar, Miriam (2001), *A migração rural no oeste paranaense/Brasil: a trajetória dos "brasiguaios"*, Scripta. Nova Revista Electrónica de Geografía y Ciencias Sociales 94/88, http://www.ub.edu/geocrit/sn-94-88.htm, last accessed 18.02.2022.

Zilles, Ana Maria Stahl/Zechlinski Maya, Leonardo/Silva, Karine Quadros da (2005), *A concordância verbal com a primeira pessoa do plural em Panambí e Porto Alegre, RS,* Organon 14/28–29, 195–219.

Zimmer, Márcia/Bandeira, Marta (2009), *A dinâmica do multilingüismo na transferência de padrões de aspiração de obstruintes iniciais entre o pomerano (L1), o português (L2) e o inglês (L3)*, https://www.academia.edu/621378/A_din%C3%A2mica_do_multiling%C3%BCismo_na_transfer%C3%AAncia_de_padr%C3%B5es_de_aspira%C3%A7%C3%A3o_de_obstruintes_iniciais_entre_o_pomerano_L1_o_portugu%C3%AAs_L2_eo_ingl%C3%AAs_L3_, last accessed 18.02.2022.

Zubaran, Maria Angélica (1994), *Os teuto-rio-grandenses, a escravidão e as alforrias*, in: Claudia Mauch/Naira Vasconcellos (edd.), *Os alemães no sul do Brasil: cultura-etnicidade-história*, Porto Alegre, Ulbra, 65–74.

Dermeval da Hora and Elisa Battisti

12 Sound-related aspects of Brazilian Portuguese

Abstract: This chapter provides a synchronic overview of sound-related aspects of Brazilian Portuguese (BP). It tackles the main characteristics of the BP sound system at the segmental and supra-segmental levels considering the heterogeneity of speech production in different BP varieties. The chapter presents the inventory of vocalic and consonantal phonemes, as well as syllable structure, word stress, intonation patterns, taking into account diatopic and diastratic variation. It also details the results of the application of various phonological processes to segments in the phonological inventory of BP: regarding vowels, the processes of vowel reduction, vowel nasalization; regarding segments in syllabic onset position, the palatalization of coronal plosives, the palatalization and vocalization of lateral approximants; regarding segments in syllabic coda position, the vocalization, velarization, rhoticization and deletion of /L/, the palatalization (*chiamento*) and deletion of /S/, the deletion (denasalization) of word-final /N/, the fricatization, retroflexion and deletion of /R/. The final part of the chapter approaches intonation patterns in BP concerning dialectal differences in melodic contours of questions and utterances as well as pragmatic related functions of prosody.

Keywords: Brazilian Portuguese, sound system, syllable, stress, phonological processes, intonation

1 Introduction

This chapter presents the main features of the phonological system of Brazilian Portuguese (BP) and the heterogeneity of its phonetic realization. Its objective is to provide a synchronic overview of sound-related aspects of BP based on the results of countless descriptive studies of the language.

The first topics in this chapter will be syllable structure and stress placement, since they affect the distribution of segments in larger units and play a role in the processes that might affect segments themselves. After this, we describe the variable behavior of segments under the influence of different vowel and consonant processes, from the perspective of variationist sociolinguistics (Labov 1972) and considering results of dialectological studies as well. Intonation will be the final topic discussed.

https://doi.org/10.1515/9783110405958-013

2 Syllable structure

Representations of the syllable structure[1] for the BP monosyllables *mas* 'but', *má* 'evil' fem., *as* 'the' fem.pl., and *a* 'the' fem.sing. are shown in footnote 1 and illustrate the internal hierarchical relations that we will refer to along this section and the chapter.

Considering the number of segments that may form syllables, as well as the nature of the segments involved, BP has syllables formed by vowels only (*a.sa* [ˈa.zɐ] 'wing'), syllables that start with one or two segments (*lu.cro* [ˈlu.kɾʊ] 'profit', *rá.dio* [ˈha.ʤjʊ] 'radio'), syllables that end in one or two segments (*car.ta* [ˈkaɾ.tɐ] 'letter', *jo.vens* [ˈʒɔ.veȷ̃s] 'youngsters', *pai* [ˈpaj] 'father'). Therefore, the shortest syllable in Portuguese contains only one segment, a vowel; and the longest contains five segments.

BP syllables may be *light* (open), with no segment in coda position—*a.to* [ˈa.tʊ] 'act', *ga.to* [ˈga.tʊ] 'cat', *pra.to* [ˈpɾa.tʊ] 'dish'; or *heavy* (closed), with one or two segments in coda position—a.*tor* [a.ˈtoɾ] 'actor', a.*trás* [a.ˈtɾas] 'behind', *mais* [ˈmajs] 'more'.

As to the segments that occupy syllable positions, BP has syllables with both simple and complex onsets. Simple onsets may be filled by any of the consonants in the system. Complex onsets may consist of a sequence of an obstruent (/p b t d k g f v/) and a liquid (/ɾ/ or /l/) consonant, for example *pr*, *pl* (*pru.mo* [ˈpɾũ.mʊ] 'level', *plu.ma* [ˈplũ.mɐ] 'feather'). There are syllables with a coda /S/ (*cos.ta* [ˈkɔs.tɐ] 'coast'), /R/ (*car.ta* [ˈkaɾ.tɐ] 'letter'), /l/ (*cal.ça* [ˈkał.sɐ] 'pants'), /N/ (*con.ta* [ˈkõⁿ.tɐ] 'account') and with a glide (*sei.ta* [ˈsej.tɐ]), but there are no syllables with plosives in the coda (the English word *club* became *clube* in Portuguese, for example). Surface realizations of the four consonantal segments that may occupy coda position in BP (/R/, /L/, /S/, /N/) are variable. Coda segments may even be deleted (see section 4.2), as in the example of coda /R/ in (01).

---

1 Phonological theory conceives syllables as being constituted by hierarchically organized units. One of the possible ways of representing the internal structure of the syllable in phonological terms is as follows (Figure 1), according to Selkirk (1982), among others.

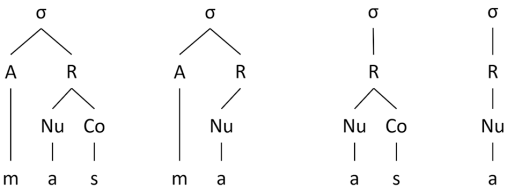


Figure 1: Internal structure of a syllable.

A syllable (σ) consists of Onset (O) and Rhyme (R). The rhyme comprises Nucleus (Nu) and Coda (Co). Any category except for Nu may be empty. Most generally vowels are mapped to Nu, consonants and glides are mapped to O and Co.

(01) /muʌɛR/ → [muˈʌɛɾ] ~ [muˈʌɛx] ~ [muˈʌɛh] ~ [muˈʌɛɽ] ~ [muˈʌɛ] ‘woman’

3 Stress pattern

According to Collischonn (2010), the following are regularities of the primary stress placement in BP:

a) primary stress is placed on one of the last three syllables of the word (three-syllables window constraint), e.g. tri*cô* ‘knitting’, *fra*co ‘weak’, *ár*vore ‘tree’, but never **á*bobora[2] ‘pumpkin’;
b) most words in Portuguese are stressed on the penultimate syllable (paroxytone) and end in a vowel. That is, when the word ends in a vowel, the stress on the penultimate syllable (paroxytone) is preferred, e.g. *ca*rro ‘car’, ma*ru*jo ‘sailor’, ca*der*no ‘notebook’, *sa*la ‘room’, *te*ma ‘theme’, es*ca*da ‘stairs’, bu*ra*co ‘hole’, es*co*la, ‘school’, *no*me ‘name’, *tro*te ‘trot’, car*du*me ‘shoal’, cordi*lhei*ra ‘range’;
c) when the last syllable of the word is closed (or heavy), stress on the last syllable (oxytone) is preferred, e.g. po*mar* ‘orchard’, fu*nil* ‘funnel’, ra*paz* ‘boy’, gar*çon* ‘waiter’, du*rex* ‘Scotch tape’. Note the following contrasts: *len*ço ‘handkerchief’—len*çol* ‘sheet’; *po*mo ‘chest’—po*mar* ‘orchard’; *rou*pa ‘clothes’—ra*paz* ‘boy’;
d) when a derivational suffix is added to the word, stress changes position, since the number of syllables in the word increases. The three-syllables window constraint leads to this relocation: *ár*vore ‘tree’—arvo*re*do ‘forest’; capi*tal* ‘capital’—capita*lis*ta ‘capitalist’; *nú*mero ‘number’—nu*mé*rico ‘numeric’. Yet adding an inflectional suffix of number (*ga*to ‘cat’—*ga*to**s** ‘cats’) does not force a change in the stress position.

As to the phonology of the primary stress, the analysis of verbs and non-verbs in BP by Bisol (1992; 1994) is of particular interest. She claims that ‘in non-verbs, stress rules operate from the non-derived word and operate again every time a new morpheme is added, during the whole derivation process, as cyclic rules; in verbs, the rules only operate when the word is ready, therefore they are noncyclic’ (Bisol 1992, 20). According to Bisol (1992), BP is sensitive to the weight of the final syllable to the right of the word. Syllable weight is relevant to the construction of metrical feet in parenthesized grids. The fact that 78 % of the words in Portuguese ending in consonants are oxytones and that the penultimate syllable of proparoxytones is not heavy support the sensitivity of BP to stress. For example, *cá*tedra ‘cathedra’, ca*dei*ra ‘chair’ are possible forms in BP, whereas **cá*deira ‘chair’ is not.

2 *Rítmico* [ˈhi.ʧi.mi.kʊ] ‘rhythmic’ seems to be one of the few cases in BP that go against the three-syllables window constraint.

The stress rule proposed by Bisol (1992, 49) is in (02).

(02) Primary stress rule
Domain: lexical word

a) Assign an asterisk (*) to the final heavy syllable, that is, a syllable with a branching rhyme.
b) In all other cases, form a binary constituent (not iteratively) with prominence to the left, of the type (* .), to the right edge of the word.

Exceptions to the rule in (02) are resolved, according to Bisol (1992), by means of extrametricality. Exceptional forms are lexically marked. In (03), there are examples of the application of the stress rule in (02) by Magalhães (2004, 42–43).

(03) Primary stress rule for non-verbs—extrametricality

a) Unmarked forms:

(* .)	(* .)	(* .)
ka.za	pa. re. de	bor.bo.le. ta
(*)	(*)	(*)
po.mar	tro.féu	co.ro.nel

b) Marked forms (proparoxytones: Ext {σ}; paroxytones: Ext {coda})

(* .)	(* .)	
fos.fo<ro>	*ar.vo<re>*	Extrametricality {σ}

(* .)	(* .)	
u.ti<l>	*vi. si. ve<l>*	Extrametricity {coda}

Oxytone words ending in a vowel, according to Bisol (1992), have an abstract final consonant, conforming to rule (3a). Evidence for such an interpretation can be found in words as *cafeC* 'coffee', *jacareC* 'alligator', *saciC* 'name of a forest myth' when the diminutive suffix "-(z)inho" is added. Such words follow the pattern of words ending in consonant: they select *-zinho* and not *-inho* as diminutive suffix (↗ 13 Morphology). For example: *pomar* 'orchard' > *pomarzinho*, *café* 'coffee' > *cafezinho*, *vovó* 'granny' > *vovozinha*, *guri* 'boy' > *gurizinho* but not *cafeinho, *vovoinha, *guriinho.

According to Bisol (1992), there is a specific rule of extrametricality for verbs, as formulated in (04):

(04) Primary stress rule for verbs—extrametricality
Mark as extrametrical:

a) the final syllable of the 1st and 2nd person plural of the imperfect forms.
Example:

(* .)	
kaN.táva<mos>	Extrametricality according to (04.a)

b) coda segments with inflectional status, that is, {N, S}. Example:
(* .)
kaN.ta<N> Extrametricality according to (04.b)

4 Segmental inventory

4.1 Vocalic system

Vowel oppositions verified in stressed syllables are neutralized in unstressed syllables in BP.

4.1.1 Stressed vowels

BP realizes a symmetrical seven-vowel system when stressed: /i, e, ɛ, a, u, o, ɔ/.[3] (↗ 4 Historical phonetics and phonology)

Chart 1: Stressed vowel system of Brazilian Portuguese.

	Coronal	Dorsal	Labial
High	i		u
Upper Mid	e		o
Lower Mid	ɛ		ɔ
Low		a	

Vowels in stressed position do not vary. BP realizes the vowel system in Chart 1 as in the examples in (05), without any dialectal or sociolinguistic variation.

(05)

mapa	[ˈmapɐ]	‘map’
tipo	[ˈʧipʊ]	‘type’
tudo	[ˈtudʊ]	‘everything’
mesa	[ˈmezɐ]	‘table’

3 The distinctive feature definition of vowel phonemes in BP in this chapter does not imply any strong commitment to the way that the distinction between upper and lower mid vowels is defined. Nevertheless, we recognize the existence of a separate aperture class node in the internal structure of vowels. See discussions by Mateus/Andrade (2000) and Wetzels (2011) on vowel height distinctions in Portuguese.

tela	[ˈtɛlɐ]	‘screen’
moço	[ˈmosʊ]	‘guy’
bola	[ˈbɔlɐ]	‘balloon’

4.1.2 Pre-stressed vowels

In BP, it is understood (Câmara Jr. 1970) that the opposition between mid vowels /e, ɛ/ and /o, ɔ/, which can be seen in stressed syllables, is neutralized in pre-stressed syllables. Compare the words *s*/a/*co* ‘bag’, *s*/ɛ/*co* (verb *secar* ‘to dry’, 1ps Pres. Ind.), *s*/*e*/*co* ‘dry’ (adjective), *s*/*i*/*co* (same as Tungiasis), *s*/*u*/*co* ‘juice’, *s*/*o*/*co* ‘punch’, *s*/ɔ/*co*, (verb *socar* ‘to punch’, 1ps Pres. Ind.) to the derived forms *beleza* ‘beauty’ from *belo* ‘beautiful’ and *ro̲chos*o ‘rocky’ from *rocha* ‘rock’, for example. The pre-stressed, first-syllable vowel in derived forms is [e] and [o] in derived forms: [b**e**'lezɐ], [r**o**'ʃozʊ].

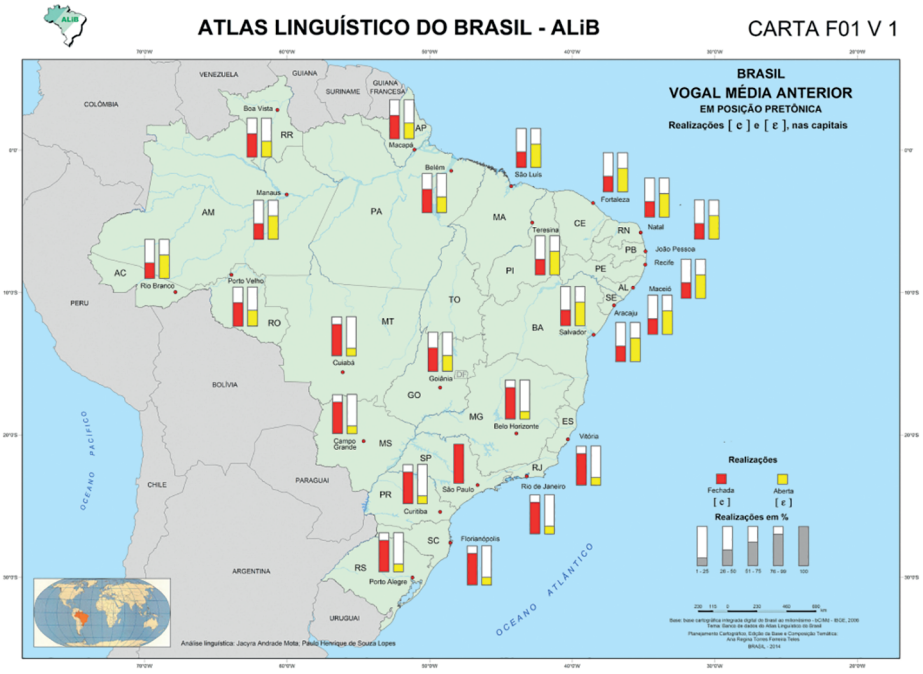
Upper mid vowels [e, o] are the most frequent pre-stressed vowels in the BP variety studied by Câmara Jr. (1970), the Cariocan dialect, spoken in Rio de Janeiro. In other varieties of BP, such as the one spoken in Bahia, lower mid vowels [ɛ ɔ] are predominant in this position. Regardless of the dialectal realizations, contrast neutralization in pre-stressed position holds. As a consequence, there is a decrease in the number of vocalic oppositions in pre-stressed position: from seven vocalic oppositions in tonic position to five in pre-stressed position.

The variable realization of pre-stressed mid vowels has aroused the interest of many scholars in Brazil, who have produced good descriptions of tendencies of use in different varieties of BP (↗ 4 Historical phonetics and phonology).

Nascentes (1922) splits Brazil into two areas according to how mid vowels are instantiated: north Brazil with open mid vowels and south Brazil with closed mid vowels.[4] The five-vowel system in pre-stressed position, as established by Câmara Jr. (1970), changes as we move from one region to the other. This has been well documented in studies by Bisol (1981) for the state of Rio Grande do Sul, Maia (1986) for Rio Grande do Norte, Silva (1989) for Bahia, Viegas (2001) for Minas Gerais, Pereira (1997) for Paraíba, Callou/Leite/Coutinho (2001) for the state of Rio de Janeiro, and Silva

4 Nascentes (1922) claims that there are two varieties of BP in the north: *amazônico* (Amazonian), comprising the states of Acre, Amazonas, Pará and part of Goiás; and *nordestino* (northeastern), comprising the states of Maranhão, Piauí, Ceará, Rio Grande do Norte, Paraíba, Pernambuco and part of Goiás. The four varieties in the south are: *baiano*, comprising Sergipe, Bahia and part of Minas Gerais and another part of Goiás; *fluminense*, comprising the states of Espírito Santo, Rio de Janeiro, Distrito Federal and another part of Minas Gerais; *mineiro*, comprising the central, western and part of the eastern regions of Minas Gerais; and *sulista* (southern), comprising São Paulo, Paraná, Santa Catarina, Rio Grande do Sul, the southern part of Minas Gerais, plus the region known as Triângulo Mineiro, as well as the south of Goiás and Mato Grosso.

(2009) for Piauí. The distribution of mid vowels in pre-stressed position is not as bilateral these days as Nascentes (1922) conceived.



Map 1: Distribution of coronal mid vowels in pre-stressed position. Source: Cardoso et al. (2014)[5].

Dialectal studies conducted by the team working on the *Atlas Linguístico do Brasil*, 'Linguistic Atlas of Brazil' (Cardoso et al. 2014; ↗ 9 Dialectology and linguistic geography), have described the distribution of coronal mid vowels and labial mid vowels in pre-stressed position. Map 1 concerns coronal mid vowels. It shows that closed coronal mid vowel [e] concentrates in the south and southeast regions. It is also present in the central-west region. It is present to a lesser extent in the north and, in the northeast, the open coronal vowel [ɛ] predominates.

The above-mentioned studies verify that: (a) in the south and southeast, there is a remarkable predominance of closed mid vowels [e] and [o] in pre-stressed position. The closed mid vowels variably raise to [i] and [u], respectively; (b) in the northeast,

5 North region: AC = Acre, RO = Rondônia, AM = Amazonas, RR = Roraima, AP = Amapá, PA = Pará, TO = Tocantins. Northeast region: MA = Maranhão, PI = Piauí, CE = Ceará, RN = Rio Grande do Norte; PB = Paraíba, PE = Pernambuco, AL = Alagoas, SE = Sergipe, BA = Bahia. Southeast region: ES = Espírito Santo, MG = Minas Gerais, RJ = Rio de Janeiro, SP = São Paulo; Central-west region: GO = Goiás, MT —Mato Grosso, MS = Mato Grosso do Sul. South region: PR = Paraná, SC = Santa Catarina, RS = Rio Grande do Sul.

central-west and northern regions, the open mid vowels [ɛ] and [ɔ] predominate, but mid vowel raising is also possible. So in the south and southwest mid and high vowels may alternate in pre-stressed position. In the other regions, three vowels may manifest in pre-stressed position: [i~e~ɛ], [u~o~ɔ], as shown in (06).

(06)

South/Southeast	North/Northeast/Central-West
[peˈzadʊ] 'heavy'	[peˈzadʊ] ~ [pɛˈzadʊ]
[moˈrɐ̃ᵑgʊ] 'strawberry'	[moˈrɐ̃ᵑgʊ] ~ [mɔˈrɐ̃ᵑgʊ]
[koˈɾuʒɐ] ~ [kuˈɾuʒɐ] 'owl'	[koˈɾuʒɐ] ~ [kɔˈɾuʒɐ] ~ [kuˈɾuʒɐ]
[meˈnĩnʊ] ~ [miˈnĩnʊ] 'boy'	[meˈnĩnʊ] ~ [mɛˈnĩnʊ] ~ [miˈnĩnʊ]

Social and structural factors condition the choice between the vowel variants. Among the structural factors, one of the most remarkable is the quality of the vowel of the stressed syllable. It may determine the realization of the pre-stressed vowel in one of two ways: either the vowel is harmonized, as in *c[o]ruja ~ c[u]ruja* 'owl', where the pre-stressed mid vowel assimilates the stressed vowel height; or it is lowered, as in *p[e]ludo ~ p[ɛ]ludo* 'hairy'. The region of the speaker conditions the raising or the lowering of the mid vowels. Speakers in the northeast tend to choose the open mid vowel, speakers in the south and southeast choose the closed mid vowel. Bisol (1981) observes that:

> '[...] one of the commonly mentioned features that distinguishes Brazilian Portuguese from European Portuguese is the pronunciation of unstressed vowels, especially pre-stressed mid vowels. /e/ is realized as [ə] or is reduced to zero, while /o/ becomes [u] more often in European Portuguese, whereas in Brazil both remain; with a possible variable switching to the high vowel' (Bisol 1981, 229).

4.1.3 Non-final post-stressed vowels

Regarding non-final post-stressed position in BP, Câmara Jr. (1970) proposes a four-vowel system resulting from the neutralization of the oppositions found in stressed position (Chart 1):

High	/i/	/u/
Mid	/e/	/.../
Low	/a/	

The non-final post-stressed vowel system does not present the open mid vowels /ɛ, ɔ/ or the closed mid vowel /o/ since neutralization favors high vowels, as in *pér[u]la* for *pér[o]la* 'pearl'.

According to Vieira (2002, 128), who analyzed speech data from the south of Brazil, non-final post-stressed vowel /o/ is never raised to [u] in some words, as in *cóc[o]ras* 'squatting' and *ânc[o]ra* 'anchor'. The behavior of the coronal vowel /e/ is different. It may be raised in some words, as in *prót[i]se* 'prosthesis', *cóc[i]ga* 'tickling', and it may be preserved in other words, as in *vésp[e]ra* 'eve', *cát[e]dra* 'cathedra'. Speech data from the northeast region, however, do not fit the proposal of Câmara Jr. (1970) neither the speech pattern verified by Vieira (2002). The tendency in northeastern BP is to prioritize open vowels. Thus, we would have *cóc[ɔ]ras* and *ânc[ɔ]ra*, for the labial; and *prót[ɛ]se* and *cóc[ɛ]ga*, *vésp[ɛ]ra* and *cát[ɛ]dra* for the coronal. The four-vowel system in post-stressed, non-final position by Câmara Jr. (1970) should be reviewed in order to capture the patterns of other regions.

4.1.4 Word-final post-stressed vowels

Word-final post-stressed vowels in BP are realized as high almost categorically (↗ 4 Historical phonetics and phonology). Examples: *rat[ʊ]* 'mouse', *leit[ɪ]* 'milk'. We are not aware of any surveys of these vowels with speech data from the north, central-west, northeast and southeast regions. Vieira (2002) focuses on the south region, analyzing data from the VARSUL[6] corpus, which covers the states of Rio Grande do Sul, Paraná, and Santa Catarina. It is worth noting that controlling the geographical distribution of the variants has played an important role in capturing the use of the mid vowels [e, o] or the high vowels [i, u], as can be seen in Table 1.

Table 1: Geographical distribution of the use of word-final post-stressed vowels per state in the south of Brazil.

	City	Application/Total	%	Relative weight
Rio Grande do Sul	Porto Alegre	48/59	81	0.99
	Panambi	10/43	23	0.29
	São Borja	33/82	40	0.44
	Flores da Cunha	21/115	18	0.22
Santa Catarina	Florianópolis	35/61	57	0.66
	Blumenau	48/77	62	0.72
	Chapecó	27/107	25	0.25
	Lages	11/44	23	0.29

6 VARSUL—Project Variação Linguística na Região Sul do Brasil is a speech data bank. See http://www.varsul.org.br/ (last accessed 18.02.2022) for information.

Table 1: (continued)

	City	Application/Total	%	Relative weight
Paraná	Curitiba	37/100	37	0.45
	Pato Branco	40/57	70	0.81
	Irati	16/76	21	0.25
	Londrina	25/56	45	0.48

Input: 0.34 Significance: 0.00
Source: Vieira (1994, 153)

Results in Table 1 express the fact that variation is significant within each state. Considering the raise of the word-final post-stressed vowel as rule application, the analysis shows that in the capital of the state of Rio Grande do Sul, Porto Alegre, vowel raising is strongly favored (0.99), whereas in Flores da Cunha, a city colonized by Italians, it tends not to apply (0.22). Results also show that in cities located in the countryside of the state of Rio Grande do Sul (Panambi, São Borja, Flores da Cunha), where BP is influenced by immigrant languages, the raising of the vowel is inhibited. Nevertheless, the tendencies are not the same in the state of Santa Catarina. Blumenau, a city colonized by Germans, has the highest relative weight (0.72), followed by Florianópolis, the capital city (0.66). In Paraná, Pato Branco exhibits the highest relative weight (0.81) while the lowest is found in Irati (0.25). Results for Curitiba, the capital city of Paraná state, and for Londrina are close to the neutral point (0.50).

The contact of BP with immigrant languages (↗ 3 The history of linguistic contact underlying Brazilian Portuguese) may have contributed to the preservation of mid vowels. This is very likely if we consider that elsewhere in Brazil, where European immigration in the 19th century was less prevalent than in the South, vowel raising is almost categorical.

4.1.5 Vowel nasalization

For historical reasons (Latin-to-Portuguese evolution) (↗ 4 Historical phonetics and phonology), vowel nasalization may be said to result from the influence of a nasal consonant following a vowel. In terms of articulation, it means that vowels are nasalized because the lowering of the soft palate, necessary for the emission of nasal consonants, begins prior to the articulation of these consonants.

In BP, the front low vowel /a/ is nasalized, raised and centralized, going from low vowel [a] to mid central vowel [ɐ̃].[7] There are no mid-open vowels or low nasalized/nasal vowels in BP.

According to Moraes/Wetzels (1992)[8] a distinction should be made between nasalized vowels and nasal vowels, that is, vowels followed by a nasal consonant in the onset of the next syllable (*ca.ma* 'bed') and vowels followed by a nasal consonant in coda position (*con.to* 'tale').

4.1.5.1 Nasalized vowels

Nasalized vowels result from the contact of the vowel with the nasal consonant that begins the following syllable, as in *c[ɐ̃].ma* 'bed', *n[õ].me* 'nome', *t[ĩ].me* 'team', *f[ũ].mo* 'smoke', *p[ɐ̃].na* 'feather' (the separation of syllables is marked by a dot). They are categorically nasalized when stressed. If unstressed, nasalization is optional, as in the right-hand column in (07), with derived forms:

(07)

banana [baˈnɐ̃nɐ] 'banana'	bananal [banaˈnaw] 'forest of banana trees'
cama [ˈkɐ̃mɐ]'bed'	camareira [kamaˈrerɐ] 'maid'
nome [ˈnõmɪ]'name'	nominal [nomiˈnaw] 'nominal'
tema [ˈtẽmɐ] 'theme'	temático [teˈmaʧikʊ] 'thematic'
rima [ˈhĩmɐ] 'rhyme'	rimado [hiˈmadʊ] 'rhymed'

The words in the left-hand column have their stressed vowel nasalized by the consonant that starts the following syllable. In the derived words in the right-hand column, the vowel that precedes the nasal is unstressed and for this reason is not nasalized. However, nasalization is variable in unstressed syllables, occurrences as *banana* [bɐ̃ˈnɐ̃nɐ] 'banana', *janela* [ʒɐ̃ˈnɛlɐ] 'window' are possible. In contrast, vowels in stressed syllables are always nasalized.

---

7 The phonetic symbol used to represent nasalized vowel /a/ is not consensual among Brazilian linguists. We follow Seara (2000, 59–60) and use [ɐ̃]. The author verifies that both stressed and unstressed nasal vowels exhibit a lower first formant (F1) than oral vowels and chooses to use the symbol [ɐ̃] to represent nasal and nasalized vowel /a/.

8 In a study of the duration of nasal vocalic segments, Moraes/Wetzels (1992) found that (a) the nasal vowel (*t**am**pa* 'cover') is longer than the oral one (*t**a**pa* 'slap'), both in stressed and pre-stressed contexts (*t**am**pa* x *t**am**pado* 'covered'); (b) the nasalized vowel (*c**a**ma* 'bed') is slightly shorter than the oral one (*c**a**da* 'each'); (c) the nasal vowel is longer than the oral one before plosives (*c**am**po* 'field') and shorter than the oral one before a fricative (*c**an**so* 'I get tired').

4.1.5.2 Nasal vowels

Nasal vowels result from the contact of the vowel with a consonant in the same syllable, as in *c[ẽ]m.po* 'field', *c[õ]n.to* 'tale', *p[ẽ]n.te* 'comb', *s[ĩ]n.to* 'I feel', *j[ũ]n.to* 'together'. These vowels undergo nasalization regardless of the tonicity of the syllable. In other words, vowels followed by a nasal consonant in the same syllable undergo nasalization be they in stressed or unstressed syllables, as in the examples in (08).

(08)

campo [ˈkẽᵐpʊ] 'field'	campestre [kẽᵐˈpɛstrɪ] 'field like'
conto [ˈkõⁿtʊ] 'tale'	contista [kõⁿˈʧistɐ] 'tale writer'
pente [ˈpẽⁿʧɪ] 'comb'	penteado [pẽⁿˈʧjadʊ] 'hair dress'
cinto [ˈsĩⁿtʊ] 'belt'	cintura [sĩⁿˈturɐ] 'waist'
junto [ˈʒũⁿtʊ]'together, close to'	juntinho [ʒũⁿˈʧĩɲʊ] 'closer'
banco [ˈbẽᵑkʊ] 'bench'	bancário [bẽᵑˈkaɾjʊ] 'bank clerk'
triunfo [triˈũᵐfʊ] 'triumph'	triunfar [triũᵐˈfaɾ] 'to triumph'

Another phonetic fact registered in the transcriptions in (08) is the presence of a transitional nasal consonant after the nasal vowel. The nasal consonant assimilates the place of articulation of the following [-continuant] consonant: the nasal is homorganic in point of articulation to the following plosive consonant. It is assumed (Câmara Jr. 1970) that the homorganic nasal consonant corresponds to a nasal archiphoneme (see section 4.2.2.4).

4.1.5.3 Final syllable

Nasal vowels can occur in the end of words. For example, *bat[õ]m* 'lipstick', *r[ũ]m* 'rum', *tamb[ẽ]m* 'also', *s[ĩ]m* 'yes', *manh[ẽ]*[9] 'morning'.

Nasal vowels are followed by a velar transitional consonant if they are back vowels /u o/ or the vowel /a/. A transitional palatal nasal follows the nasal vowel after the front vowels /i e/. A bilabial transitional nasal is never realized, despite the suggestion of the spelling conventions. Nasal vowels /e, o/ are variably diphthongized. Examples are in (09).

9 The nasal segment in the end of words like *lã* 'wool', *romã* 'pomegranate' may be absent from the written form, but we assume it is present in the phonological representation of the word.

(09)

batom [baˈtõⁿ]~[baˈtõw̃ⁿ]	‘lipstick’
também [tẽᵐˈbẽɲ]~[tẽmˈbẽj̃ⁿ]	‘too’
fim [ˈfĩⁿ]	‘end’
atum [aˈtũⁿ]	‘tuna fish’
galã [gaˈlẽⁿ]	‘handsome man’

4.1.5.4 Nasal diphthongs

Nasal diphthongs as in *pão* ‘bread’, *põe* ‘he/she puts’, *mãe* ‘mother’ are assumed (Williams 1968) to be remnants of the transformation of Latin into Portuguese. Today, the most productive nasal diphthong, that is, the one most frequently found in new words is the diphthong *ão*, plural form *ões*.

Although not written, a transitional nasal consonant may be produced after the semivowel. A word like mãe ‘mother’, for example, may be produced as [ˈmẽj̃] or [ˈmẽj̃ⁿ]. The presence of the transitional nasal is attested when, for example, children call out their mothers, shouting “manhe” [ˈmẽɲe].

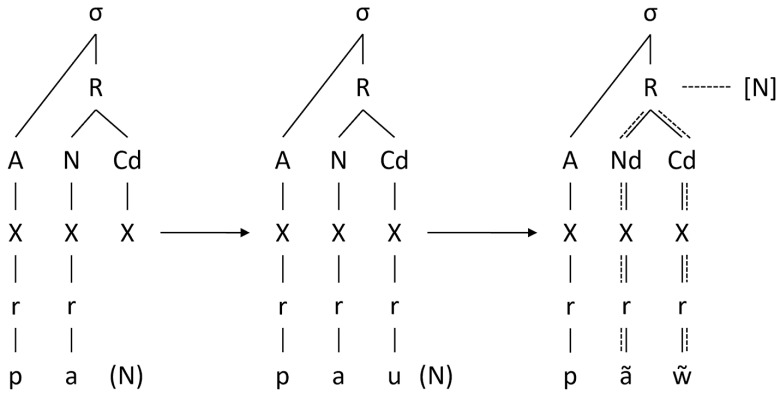
There are different approaches regarding the phonological representation of the nasal diphthong. Câmara Jr. (1970, 50) considers it to be a sequence of two vowels plus a nasal archiphoneme, where the glide derives from the theme vowel: /auN/, for example, with /N/ in coda position. However, this idea can be questioned in that the syllable pattern of Portuguese only admits /S/ in the second position of a complex coda.

Wetzels (1997, 222–227) claims that nasal diphthongs in noun forms such as *irmão* ‘brother’, *mãe* ‘mother’ and *muito* ‘much/many’ are lexicalized, i.e., they are not derived from an underlying sequence of vowel plus nasal consonant. He also admits a certain degree of lexicalization in nasal diphthongs of inflected verb forms: in *falavam* ‘they used to speak’, for example, the nasality of the theme vowel or that of the morpheme marking the imperfect originates from the inflectional suffix, which the author assumes to be an underlying /ũ/.

Bisol (1998) assumes two processes of nasalization. She considers all nasal diphthongs as derived forms, except those within words, as in *caimbra* ‘cramp’ and *muito* ‘much/many’. These forms are lexicalized due to their exceptional nature. The processes are: (a) assimilation, where N expands to the vowel and gains articulation features of the following consonant or of the vowel nasalized by it; (b) stabilization, as the rhyme joins a floating nasal, which does not acquire articulation features. The former creates the internal nasal vowel (*canto*, *senda*, *samba*) or the external diphthong of variable nature in words without a theme vowel (*homem*, *fórum*, *jardim*); word-finally, there is mutual assimilation: N nasalizes the vowel, and the vowel creates the homorganic glide. The latter creates a true nasal diphthong in words with a theme vowel: *irmão*, *põe*, *pão*. N is dissociated since it does not acquire articulation features and, due to the effects of stability (Goldsmith 1976) it remains floating. The theme

vowel takes the position of the syllable structure left empty by N, which is reassociated with the rhyme and percolates up to the terminal segments. A general rule converts a mid vowel adjacent to another vowel into a high vowel, and a universal rule that creates diphthongs forms the glide. The representation of the derivation of a nasal diphthong according to Bisol (1998) is set out in (10).

(10)



Source: Bisol (1998, 5), where (N) means floating nasal.

It is worth noting the existence of other analyses on the same theme (Parkinson 1983; Magalhães 1990; Girelli 1988; Battisti 1997). The phonological representation of nasal diphthongs remains an open issue.

4.2 Consonantal system

Some consonants in onset and coda position exhibit alophony.

4.2.1 Consonants in onset

The system of consonants in onset in BP is in Chart 2.

Chart 2: Onset consonants of Brazilian Portuguese

	Labial	Coronal		Dorsal
		+ anterior	- anterior	
Stops	p b	t d		k g
Fricatives	f v	s z	ʃ ʒ	
Rhotics		r		
Laterals		l	ʎ	
Nasals	m	n	ɲ	

4.2.1.1 Coronal plosive consonants

Coronal plosive consonants /t, d/ in BP are variably palatalized, becoming affricates [ʧ] and [ʤ] as the result of regressive assimilation triggered by the following vowel, a coronal high or a coronal mid vowel that is raised, as the examples in (11) show.

(11)

Contexts of coronal high vowels	Contexts of raised coronal mid vowels
[di]tado ~ [ʤi]tado 'dictation'	po[te] > po[ti] ~ po[ʧi] 'pot'
re[ti]ro ~ re[ʧi]ro 'retreat'	on[de] > on[di] ~ on[ʤi] 'where'
Po[ti] ~ Po[ʧi] 'Poti' (female name)	

A palatal glide also triggers regressive palatalization: *rá[ʤj]o*, 'radio'. Palatalization is very common in most Brazilian regions, mainly in urban areas, and has been the object of study of several researchers.[10] Bisol (1985) investigated the palatalization of /t, d/ in the state of Rio Grande do Sul. She analyzed speech data from four groups of speakers: 15 monolingual speakers of BP from Porto Alegre, 15 BP speakers from the frontier area, where BP is in contact with Spanish (↗ 11 Brazilian Portuguese: contemporary language contacts), 15 bilinguals (BP-German dialects) from the German colonization area, 15 bilinguals (BP-Italian dialects) from the Italian colonization area (↗ 3 The history of linguistic contact underlying Brazilian Portuguese). All these speakers attended elementary school only. A control group was used, comprising 15 BP speakers from Porto Alegre who had a university diploma.

10 Lopez (1979) analyzed the dialect of the city of Rio de Janeiro and Bisol (1985) worked with speech data collected in communities in the state of Rio Grande do Sul. More recent studies were carried out with data collected in Rio Grande do Sul (Almeida 2000; Dutra 2007; Battisti et al. 2007). Hora (1990) analyzed the process in Alagoinhas (state of Bahia), Macedo (2004) worked with data from communities in the city of Rio de Janeiro, and Pagotto (2004) analyzed speech data from Florianópolis, capital city of Santa Catarina state.

Bisol treated the speech data quantitatively. The following structural factors were selected in the analysis: preceding and following context, syllable and liaison. Ethnicity was the only social factor selected. As to preceding and following contexts, the study highlighted the important role of the sibilant consonant, creating alternations such as: *in[sti]tuto~in[stʃi]tuto* 'institute', *pare[dis]~pare[dʒis]* 'walls', disfavoring palatalization. Concerning syllable structure, it was observed that palatalization occurs in closed and open syllables, but obeys the following hierarchy: stressed, pre-stressed and post-stressed, in examples such as *an[tʃi]go* 'ancient', *[dʒi]lema* 'dilemma' and *vin[tʃi]* 'twenty', respectively. Regarding liaison, the word-initial coronal plosive is likely to be preserved: *[d]escabelado* 'with untidy hair'. Also concerning liaison, the role of the prefix and the clitic were analyzed. Negative prefixes *de-*, *des-* and *dis-* refrain palatalization. Unlike these, the clitics *de* 'of' and *te* 'you, object pronoun' are sensitive to the rule.

In terms of ethnicity (the only social factor selected), Bisol (1985) suggests that the dialect of the state of Rio Grande do Sul due to contact with immigrant languages (↗ 3 The history of linguistic contact underlying Brazilian Portuguese) that do not favor palatalization, such as Italian dialects has hindered the expansion of the rule.[11] One important finding of the study is that palatalization of the coronal plosives in Rio Grande do Sul is constrained by a sibilant following the triggering vowel, as in *disputa* 'dispute', because the triggering vowel is deleted in this context.

The study by Hora (1990) used speech data collected in Alagoinhas, a small city in Bahia, a state located in the northeast region of Brazil. The sample was stratified by social class, gender, age and style. The following structural variables were controlled: following and preceding phonological contexts, sonority, tonicity, syllable position.

Five factor groups were selected in the statistical analysis: following and preceding phonological contexts, sonority, tonicity and syllable position. Two of these, the following and preceding phonological contexts, were the same as those selected in Bisol's study, confirming that structural factors play an important role in the conditioning of the process. Hora (1990) concludes that palatalization tends to occur in the speech of the upper and middle class subjects. It is favored by 15 to 47-years-old speakers, regardless of gender and it is peculiar to formal styles of language use, such as reading. He also concludes that palatalized variants are prestige forms since they are more frequently used in formal styles and by those in higher social classes and with higher levels of education.

Both studies on regressive palatalization (Bisol 1985; Hora 1990), one conducted in the south and the other in the northeast of Brazil, cover a variable process that is observed throughout the country and that is subject to social and structural restric-

11 Recent analyses such as the one by Battisti/Dornelles Filho (2015) show that the influence of the contact of BP with immigrant languages has decreased and palatalization in BP has progressed in the Italian immigration area.

tions. Considering the differences in the communities studied and the stratified profiles of the subjects, one might note that the conditioning factors of rule application are rather similar. The underlying coronal high vowel is the major trigger.

Progressive palatalization of /t, d/ can also be verified in BP, but it is not as widespread as the regressive process. It is found only in speech communities in the northeast of the country, such as João Pessoa, in the state of Paraíba. Examples of this kind of palatalization are given in (12).

(12)

oito	[ˈojtʊ] ~ [ˈotʃʊ]	'eight'
jeito	[ˈʒejtʊ] ~ [ˈʒeʧʊ]	'way'
leitura	[lejˈturɐ] ~ [leˈʧurɐ]	'reading'
doida	[ˈdojdɐ] ~ [ˈdoʤɐ]	'crazy, fem.'
gosto	[ˈgoʃtʊ] ~ [ˈgoʃʧʊ]	'taste'

The trigger of progressive palatalization is not the subsequent vowel. In the examples in (12), vowels immediately after the coronal plosives are not coronal high vowels neither the palatal glide. Even when a possible trigger (of regressive palatalization) is present, the process is inhibited, as in *noite* 'night', whose pronunciation is ['nojti] in the BP spoken in the northeast of Brazil. The variable rule applies in the presence of a palatal glide in the preceding context, which triggers the process of progressive palatalization. It is worth mentioning that progressive palatalization only applies if the vowel after the plosive is a labial or dorsal, as in *jeito* [ˈʒeʧʊ] 'way', *coitado* [koˈʧadʊ] 'pitiful'. Another trigger of progressive palatalization is preceding palatal fricative consonant, as in *gosto* [ˈgoʃtʊ]~[ˈgoʃʧʊ] 'taste'.

4.2.1.2 Lateral approximants

The lateral approximant verified in syllabic onset in BP is generally the alveolar one, both word-initially (as in *lado* 'side', *leve* 'soft', *lodo* 'mud', *luta* 'fight', *lima* 'lime') and word-internally (as in *alado* 'winged', *elevador* 'elevator', *tolo* 'fool', *aluno* 'student', *apelido* 'nickname').

In northern Brazil, in the states of Pará and Amapá, however, when the alveolar lateral is followed by /i/, it may be variably realized as palatal, as in *[l]ivro~[ʎ]ivro* 'book', *a[l]i~a[ʎ]i* 'there'. In Belém, capital city of Pará, the palatal lateral approximant is categorical in such an environment, according to Oliveira/Lima/Razky (2009; 2016).

The palatal lateral approximant in syllabic onset is rare in BP in word-initial position. It is found word-internally. Freire (2016) shows that in the Paraiban variety of BP the palatal lateral is sometimes realized as an alveolar lateral or as a palatal glide: *mu[ʎ]er~mu[l]er~mu[j]er* 'woman'. In most cases, however, the alternation occurs between the palatal and the glide, as in *traba[ʎ]o~traba[j]o*.

4.2.2 Consonants in coda

Coda of VC and CVC syllables in BP can be filled by /R, L, S, N/ only (Câmara Jr. 1970).

Chart 3: Possible codas in Brazilian Portuguese.

/L/		/R/		/S/		/N/	
mid	final	mid	Final	mid	final	mid	final
fal.ta 'lack' al.to 'high'	pa.pel 'paper'	car.ta 'letter' ar.te 'art'	tu.mor 'tumor'	pas.ta 'folder' es.ti.lo 'style'	mas 'but'	cam.po 'field' an.jo 'angel'	bom 'good'

Except for /S/, all coda segments are sonorants in BP. Foreign words with different kinds of consonants in coda position form a new syllable with an epenthetic vowel when incorporated into BP: *club* > *clube* ['klubɪ] 'club'.

4.2.2.1 Rhotics

Rhotics in BP are extremely variable. There are multiple variants in coda position, as in the examples *carta* 'letter' and *garfo* 'fork' in (13.a), *mar* 'sea' and *cantar* 'to sing' in (13.b).

(13)

	[r]	[r]	[x]	[ɾ]	[h]	[ø][12]	glide
a.	ca[r].ta ga[r].fo	ca[r].ta ga[r].fo	ca[x].ta ga[x].fo	ca[ɾ].ta ga[ɾ].fo	ca[h].ta ga[h]fo	–––– ga[ø].fo	ca[j].ta ga[w].fo
b.	ma[r] can.ta[r]	ma[r] can.ta[r]	ma[x] can.ta[x]	ma[ɾ] canta[ɾ]	ma[h] can.ta[h]	ma[ø] can.ta[ø]	–––– ––––

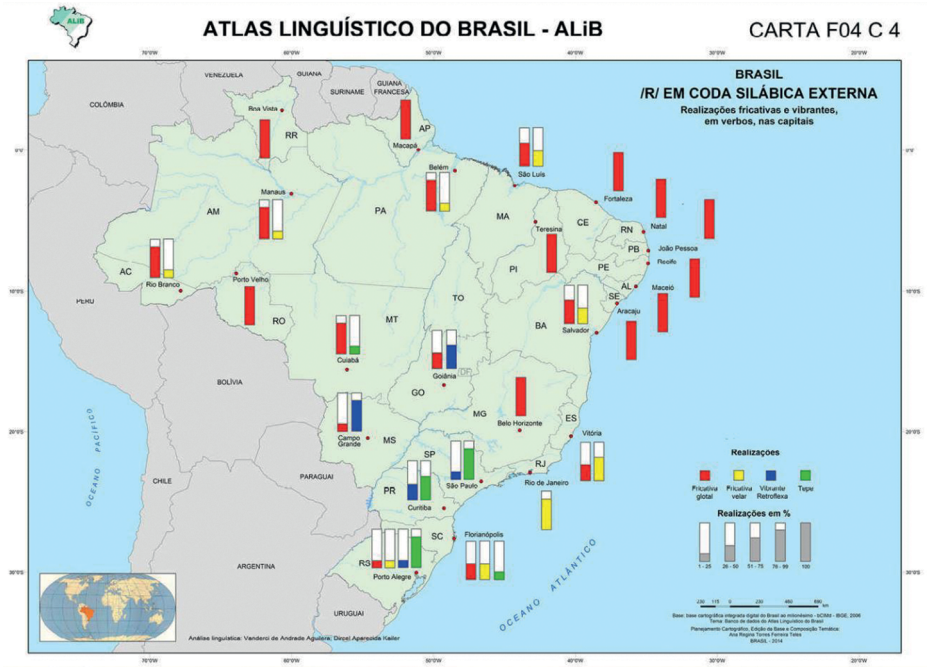
Brazilian dialects select a subset of coda variants from the inventory in (13). In BP spoken in northeastern communities as João Pessoa, Paraíba, coda variants are [ø] and [h]. In word-final syllables, [ø] is more productive than [h]. In word internal syllables, [ø] is observed before a fricative, as in the examples in (14), [h] before other consonants.

12 Symbol [ø] represents deletion.

(14)

força [ˈfoøsɐ]	‘strength’
várzea [ˈvaøzɛɐ]	‘lowland’
garfo [ˈgaøfʊ]	‘fork’
cerveja [seøˈveʒɐ]	‘beer’
marcha [ˈmaøʃɐ]	‘march’
gorjeta [goøˈʒetɐ]	‘tip’

Among rhotics, /ɾ/ and /r/ are the only contrastive units in BP. Contrast is observed in intervocalic position: ca/ɾ/o ‘expensive’, ca/r/o ‘car’. In onset position, both word-initially and in intervocalic position, /r/ is realized more frequently as [x] or [h], rarely as [r]. /ɾ/ does not occur in the onset of word-initial syllables.



Map 2: Distribution of /ɾ/ in word-external syllable coda in BP spoken in capitals of Brazilian states. Source: Cardoso et al. (2014).

Monaretto (1992) showed that in the speech of Rio Grande do Sul the deletion of rhotics in word-final position is more productive in verbs than in nouns. In João Pessoa and Rio de Janeiro, according to Votre (1978) and Callou/Moraes/Leite (1994), deletion occurs both in verbs and in nouns.

Map 2 provides an overview of the distribution of four frequent variants of /R/ in word-final coda of infinitives in capital cities of Brazilian states. A relatively produc-

tive variant of /R/ is the retroflex approximant [ɻ] (*caipira* 'r'), although it is geographically limited to interior areas of the states of São Paulo, Paraná and Mato Grosso do Sul, and to the southern area of the states of Goiás, Mato Grosso and Minas Gerais (Aguilera 2011, 125). In this broad region, the retroflex variant occurs both word-internally and word-finally.

In sequences of words, word-final rhotics may be resyllabified if there is a vowel in the beginning of the following word. They become the onset of the resulting syllable, as in *mar abaixo* > *ma.ra.bai.xo* 'sea below'.

4.2.2.2 Laterals

In BP, [w], [ɬ] and [ø] are the most productive variants of the lateral consonant in coda position. Other possible variants exist. In popular speech, for example, a lateral can alternate with a glottal fricative resulting from a rhotic, as in *fa[w]ta ~ fa[h]ta (< /farta/)* 'absence'. Examples of variants [w], [ɬ], [ø] are in (15).

(15)

Word-internal coda			Word-final coda		
[w]	[ɬ]	[ø]	[w]	[ɬ]	[ø]
pa[w]co 'stage'	pa[ɬ]co	----	jorna[w] 'newspaper'	jorna[ɬ]	jorna[ø]
de[w].ta 'delta'	de[ɬ]ta	----	pape[w] 'paper'	pape[ɬ]	pape[ø]
bi[w]tre 'rotter'	bi[ɬ]tre	----	ani[w] 'indigo'	ani[ɬ]	ani[ø]
fo[w]ga 'clearance'	fo[ɬ]ga	fo[ø]ga	so[w] 'sun'	so[ɬ]	so[ø]
co[w].cha 'bedspread'	co[ɬ].cha	co[ø]cha	azu[w] 'blue'	azu[ɬ]	azu[ø]
cu[w]pa 'guilt'	cu[ɬ]pa	cu[ø]pa			

Variant [w] in coda position, both word-internally and word-finally, is the most recurrent in BP. It is found from north to south of Brazil, and its use does not depend on speakers' gender, age or level of education.

Vocalization of the lateral consonant in syllabic coda has important implications in writing. It is very common to see the substitution of letter "l" for letter "u", mainly word-finally, since such forms exist in written BP, as in *degrau* [deˈgraw] 'step', *véu* [ˈvɛw] 'veil'. A study with elementary school students (Tasca 2002) shows that, mainly

with new words, there is a strong tendency for the substitution of the grapheme <l> by <u>, as is *jovial* [jovi'aw] 'jovial'.

The occurrence of the velarized alveolar variant [ɫ] in syllabic coda is strongly associated with the age of the speaker. Studies of BP spoken in the south of Brazil (Quednau 1993; Tasca 1999; Espiga 2003) verified coda [ɫ] mainly in communities in the interior of the state of Rio Grande do Sul in the speech of elderly people. By contrast, in a study by Hora (2006) in João Pessoa, in the northeast of Brazil, [ɫ] is not very productive. When it does occur, age is the major conditioning factor: older speakers are those who favor coda [ɫ] the most.

In word-internal codas, deletion of the vocalized coda never occurs if the vowel that precedes the lateral is coronal or low, as one sees in (15). If the vowel is labial, there is a sort of grading towards a raised segment and, as the vowel raises, deletion becomes more predictable. The direction, then, would be [ɔw] > [ow] > [uw] > [u], as in *v[ɔw]ta* 'return', *t[ow]do* 'awning' and *c[uw]pa~c[u:]pa* 'guilt'.

In word-final position, deletion seems to be differently conditioned. Its realization is related to the speaker's level of *education*. In general, speakers with less education are more likely to delete the lateral consonant, except when the preceding vowel is /u/, as in *azul* 'blue', according to data from the city of João Pessoa (Hora 2006).

Studies on BP spoken in the state of Rio Grande do Sul found no evidence of the deletion of the lateral in any position. Quednau (1993) analyzed vocalization of the lateral in syllabic coda in BP in monolingual (BP) and bilingual (BP and immigrant languages—German and Italian dialects) communities considering the variants [ɫ] and [w]. Tasca (1999) investigated the realization of the lateral in coda position in communities where BP is in contact with immigrant languages (↗ 3 The history of linguistic contact underlying Brazilian Portuguese) and controled [l] and [ɫ]. Espiga (2003) studied the coda lateral in data collected in communities where BP is in contact with Spanish (↗ 11 Brazilian Portuguese: contemporary language contacts) and considered variants [l] and [lʷ].

So, surveys carried out in Brazil on the lateral consonant in coda position allow us to conclude that the vocalized form is the most productive among all variants. The other variants are conditioned by the age of speakers, as in the case of [ɫ], or by their level of education, as in the case of [ø].

4.2.2.3 Coronal fricatives

Scholars such as Teyssier (1980), Mattos e Silva (1996) and Cardeira (2006) agree that the first reference to palatalization of /S/ in syllabic coda (named *chiamento*) is found in Verney (1746) (↗ 4 Historical phonetics and phonology). The most likely hypothesis is that syllable-final /*s*/ and /*z*/ pronounced as dental-alveolar became palatal in European Portuguese (EP) around the 16th century. In BP, palatalization of coda /S/ is variable. In BP varieties characterized by the process, voiceless palatal [ʃ] occurs before voiceless consonants and in word-final position, as in *faz frio* 'it is cold', *vista*

'sight' and *atrás* 'behind'. The voiced palatal [ʒ] occurs before the voiced consonant, as in *mesmo* 'same' and *depois dele* 'after him'.

Considered a common feature of contemporary EP by Teyssier (1980), palatalization of /S/ in coda position did not become general in northern dialects of EP or in Galician or in BP, as mentioned before. In dialects of northern Portugal, coda *-s* and *-z* are generally realized as apical-alveolar segments. In some of them, *-s* is pronounced as an apical-alveolar fricative, *-z* is realized as a pre-dorsal-dental fricative.

In BP, alveolar is the variant of /S/ more generally verified. The palatal-alveolar pronunciation is characteristic of certain speech communities, such as Rio de Janeiro and Belém.

Like rhotics, /S/ in syllabic coda in BP has been the object of numerous studies, using data from every Brazilian regions or from specific regions. Attested variants of coda /S/ are [s], [ʃ], [z], [ʒ], [h], [ø]. Examples are in (16).

(16)

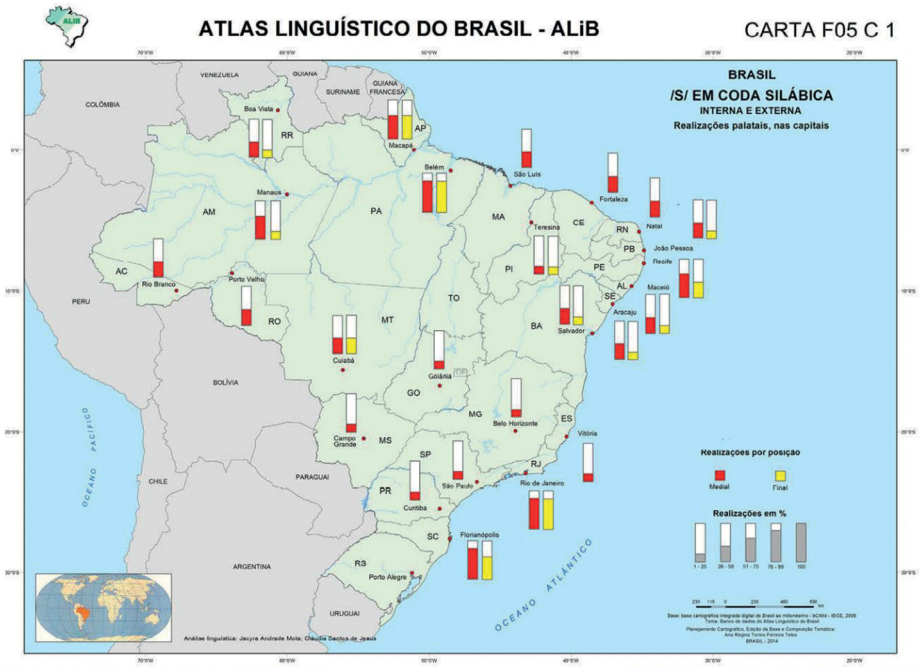
Variants	Word-internal coda	
[s]~[ʃ]	*casca* [ˈkaskɐ]~[ˈkaʃkɐ]	'peel'
[z]~[ʒ]~[h]	*desde* [ˈdezdʒɪ]~[ˈdeʒdʒɪ]~[ˈdehdʒi]	'since'
[ø]	*mesmo* [ˈmeømʊ]	'same'
[s]~[ʃ]~[h]	Word-final coda *lápis* [ˈlapis]~[ˈlapiʃ]~[ˈlapiø]	'pencil'

Variationist sociolinguistic studies (Labov 1972) of /S/ in coda position in BP allow us to sketch the distribution of coronal fricative variants in different speech communities. In word-internal position, the most productive among the six variants of coda /S/ are the alveolars [s, z] and the palatal-alveolar [ʃ, ʒ] ones. The alveolars occur in most Brazilian speech communities. The study by Callou/Moraes/Leite (1994) using data from NURC *corpus*[13] shows that alveolars are preferred in the cities of Porto Alegre, São Paulo and Salvador, but not in Rio de Janeiro and Recife. Hora (2003) verifies that alveolars are preferred in Paraíba state, noting however that the palatal-alveolars may also occur if the following context is [t] or [d], as in *leste* le[ʃ]te 'east' and *desde* de [ʒ]de 'since'.

13 NURC—Project *Norma Urbana Culta*, 'Urban Educated Norm'. NURC is a speech data bank. See http://www.nurcrj.letras.ufrj.br/ (last accessed 18.02.2022) and https://fale.ufal.br/projeto/nurcdigital/ (last accessed 18.02.2022) for information.

In word- final position, unvoiced coronal fricatives [s, ʃ] are preferred in BP. Deletion of the word-final fricative is rare, both when it is part of the lexical morpheme (*menos* 'minus') and when it is a mark of number (*casas* 'houses').

Dialectal studies for the *Atlas Linguístico do Brasil* (Cardoso et al. 2014) show (Map 3) that palatal variants of coda /S/ in word-internal position are more productive in Florianópolis, in the south of Brazil, Rio de Janeiro, in the southeast, Recife, in the northeast, and Belém, Macapá and Manaus, in the north. This finding confirms the results of most sociolinguistic studies, which show that the alveolar fricative is the most frequent variant of /S/ in syllabic coda in BP.



Map 3: Distribution of /S/ in syllable coda in the state capitals of Brazil. Source: Cardoso et al. (2014).

In a study with data collected in Rio de Janeiro city, Guy (1981) proposes two separate rules to account for the absence of the word-final fricative. One rule is phonological, it targets lexical *-s* (*menos* 'less', *lápis* 'pencil). The other rule concerns the plural suffix *-s*. The author says that in noun phrases in BP not all words are suffixed with -s. For example, the noun phrase *a casa grande* 'the big house' may contain different plural markings: *as casas grandes*, *as casas grande*, *as casa grande* 'the big houses'. In the example, definite article *as* is the only form always carrying plural suffix *-s* (↗ 3 The history of linguistic contact underlying Brazilian Portuguese).

Deletion of word-final coda /S/ is sensitive to the position of primary stress, such that /S/ is more resistant to deletion in stressed syllables than in unstressed ones. For

example, /S/ is more easily deleted in words like *ônibus ~ ônibu* 'bus', *lapis ~ lápi* 'pencil' than in words like *rapaz* 'boy', with final stress. Moreover, deletion of /S/ is less frequent in female speech than in male speech.

4.2.2.4 Nasals

As seen above, post-vocalic nasal consonants in word- internal position nasalize preceding vowels in BP and may be phonetically realized as transitional segments with the place of articulation of the following stop or preceding vowel (Cagliari 1977): *junto* [ˈʒũⁿtʊ] or [ˈʒũⁿtʊ]. In word-final position, they can be deleted in unstressed syllables: *homem* ['õmẽⁿ]~['õmɪ] 'man'. The different possibilities of realization of the nasal consonant in coda position, determined by context, evidence the hypothesis of Câmara Jr. (1970) that a nasal archiphoneme is the phonological representation of coda nasals in BP.

The deletion process that applies to coda /N/ in word-final, unstressed syllable (as in *homem* 'man') applies to unstressed diphthong *-ão*,[14] as in *órfão>órfu* 'orphan'. Denasalization is variable in BP and the process has received the attention of some scholars (Guy 1981; Battisti 2002; Schwindt/Bopp da Silva 2010).

Guy (1981) confirms that denasalization does not apply to stressed final syllables. The author analyzed the speech of twenty informants from Rio de Janeiro, nine women and eleven men, with ages ranging from fifteen to fifty-four years old, all students at MOBRAL, a program for late literacy. Guy (1981) verifies that preceding, adjacent nasal consonants (*co**m**em* 'they eat') or nasal consonants in the beginning of the following word (*comem **m**uito* 'they eat too much') inhibit the deletion of a final nasal. Preceding, adjacent velars (*bri**g**am* 'they fight') and palatals (*su**j**am* 'they smudge') favor the process. Regarding social variables, Sex is the only variable that affects denasalization: men denasalize more than women. Age does not exhibit clear effects, so Guy (1981) cannot say that the variable denasalization is change in progress in the speech community investigated.

Battisti (2002) studies variable denasalization in BP spoken in the south of Brazil. The analysis of speech data from ninety informants of the VARSUL data bank, from three different states—Rio Grande do Sul, Santa Catarina, Paraná—confirmed the tendency observed by Guy (1981): denasalization seems to be stable (not in progress) in BP. The total proportion of rule application is 43 %. Denasalization is favored by informants from Santa Catarina and is conditioned by nouns and by vowels in the beginning of the following word.

Schwindt/Bopp da Silva (2010) also investigate denasalization in BP spoken in the south of Brazil. They enlarge Battisti (2002) sample and include other factors in the

14 The nasal diphthong *-ão* is stressed in most words (*balão* 'ballon', *avião* 'airplane', *multidão* 'crowd', *limão* 'lemon', *irmão* 'brother').

analysis, such as Tonicity of the following context. They investigate the speech of 12 communities and verify that the total proportion of rule application is 34 %. Denasalization is favored by young people with a low level of education and is conditioned by vowels in the following context. Word class affects denasalization: verbs inhibit the process. This finding suggests that morphology has something to do with denasalization.

5 Intonation—Intonation patterns

Studies on intonation in BP have emerged only recently and are far from being general or conclusive (↗ 4 Historical phonetics and phonology). We will refer initially to those by Moraes (1993; 1998) with the results of detailed acoustic analysis of data on the variety of BP spoken in Rio de Janeiro in its cultured, educated form.

Moraes investigates the modality of utterances. Some relevant findings of his analyses are:

(a) lexical stress is basically manifested in acoustic terms through the conjunction of the parameters duration and intensity;
(b) phrase or sentence accents are the phonetic expression of groupings of words into syntactical informational units;
(c) accent is located on the lexical item farthest to the right in the prosodic group (on the tonic syllable) and is a supplementary indication of the acoustic projection of lexical stress in this context;
(d) the physical correlate of accent is pitch (fundamental frequency or F_0);
(e) the melodic level on certain syllables of the sentence, especially on the final tonic, defines whether the utterance is declarative or interrogative, for example.

The neutral declarative pattern in BP is the same as in most known languages: it is characterized by a drop in fundamental frequency (F_0) at the end of the utterance, on the final tonic, while the initial pitch is at a mid-level (Moraes 1998). For example, in an utterance such as *Já foi* 'He has already left', which is a possible answer to the question *E o José?* 'And what about José?', the mid-level falls on the first syllable *já* while the drop occurs on the following syllable *foi*. Longer declarative utterances present overall behavior of a moderate, continuous drop in pitch over the entire sentence, especially over the unstressed syllables.

Silvestre/Cunha (2013) analyze the intonation of neutral assertive utterances in three Brazilian capital cities: Natal (northeast region), Rio de Janeiro (southeast region) and Porto Alegre (south region). They use speech data from ALiB, Atlas Linguístico do Brasil (Cardoso et al. 2014) and perform acoustic analysis on sixty recordings of utterances such as *Morreu eletrocutado* 'He has died electrocuted', *O sol tá nascendo* 'Sun is rising', *Manda pelo correio* 'Send by mail'. The authors verify three different regional patterns for neutral assertive utterances: (a) prominent F_0 at the pre-nuclear

stress in Natal, (b) similar levels of F_0 at the pre-nuclear stress and at the nuclear stress in the intonational phrases of speakers from Rio de Janeiro, (c) prominent F_0 at the nuclear stress in the intonational phrases of speakers from Porto Alegre. Social factors such as age, sex and level of education do not affect the intonation of neutral assertive utterances.

Intonation is the main device for signaling questions in BP. Average pitches are higher in interrogative than in declarative utterances, as the great majority of languages (Moraes 1998). The shape of the interrogative melodic contour of yes/no questions is different from that of WH-questions in BP. Yes/no questions display a pitch rise on the last accented syllable. Comparing yes/no questions with declarative utterances, Moraes (1998) notes that: a) the initial pitch of the yes/no questions is slightly higher than in declarative utterances; b) the final pre-tonic of a yes/no question is lower than the final pre-tonic of a declarative utterance, thus creating a greater contrast with the final tonic. While descending in a question, any existing post-tonic syllable(s) will be even lower in a declarative utterance.

In WH-questions that begin with a question word (*Quando você soube?* 'When did you find out?'), the rise occurs on the first accented syllable of the utterance. However, when the interrogative particle is placed at the end of the utterance (*Você soube quando?* 'You found out when?'), there is a rise in pitch on the accented syllable immediately preceding it. This pattern is also used in alternative questions. Here, a higher pitch rise on the accented syllable preceding the alternative particle "ou" informs that the utterance is a question not a statement: *Ele quer doce ou fruta?* 'Does he want sweets or fruit?'

Interrogative patterns vary in BP. A frequent one is the double-rise pattern in yes/no questions, where the first rise occurs on the first accented syllable and the second—a smaller one—on the final accented syllable. This pattern appears in many different situations, including rhetorical yes/no questions, questions introduced with *Será que* (I wonder / Do you think), and requests: *Será que vai chover?* 'I wonder if it will rain?'. The double rise, yes/no question pattern is frequently used to start up a conversation: the question is delivered to the listener as new information.

Bernardo da Silva (2011) investigates the intonation pattern of yes/no questions in BP. The author uses speech data from ALiB. She performs acoustic analysis on recordings of speakers from twenty-five of the twenty-seven Brazilian capital cities, with questions like *Você vai sair hoje?* 'Are you going to go out today?'. Bernardo da Silva (2011) verifies that the main difference in regional dialects concerning the intonation of yes-no questions lies on nuclear stress: a rising movement prevails in the speech of northeastern capital cities and a circumflex configuration characterizes the speech of all other four Brazilian regions (north, central-west, southeast and south). The alignment of the peak on the last stressed syllable of the utterance is typical of the speech of southeastern capital cities.

Rosignoli/Fernandes-Svartman (2016) follow Intonational Phonology (Pierrehumbert 1980; Pierrehumbert/Beckman 1988; Ladd 1996; [2]2008; Jun 2005) in their study of

the intonational patterns of global interrogative sentences in Paulista dialect. They investigate prosodic distinctions and similarities that can define nine types of global interrogative sentences: Global neutral-Information seeking (*O João vai à festa*? 'Does John go to the party?'), Global-Neutral-Disjunctive (*Você vai ao mercado ou à farmácia?* 'Are you going to the market or to the drugstore?'), Global-Neutral-Rethorical (*Será que eu fiz a melhor escolha?* 'Did I make the best choice?'), Global-Non-Neutral-Focalized (*É amanhã que você vai à entrevista?* 'Is it tomorrow that you are going to the interview?'), Global-Non-Neutral-Negative (*Você não vai estudar?* 'Aren't you going to study?'), Global-Non-Neutral-Confirmative (*É aqui mesmo o local da festa?* 'Is the party really here?'), Global-Non-Neutral-Echo Manipulative (*Entendeu?*/ *Concorda?* 'Did you understand?/Do you agree?'), Global-Non-Neutral-Echo Absolute (O quê?/ Como disse? 'What? How did you say?'), Global-Non-Neutral-Incredulity (*Ele abandonou o filho?!* 'Did he abandon his child?!'). The analysis reveals that both nuclear tonal accent and boundary tone are indispensable for the characterization of the sentence types. For example, the non-neutral global interrogative of the echo type is distinguished by a high boundary tone; total contour is crucial to distinguish rhetorical neutral global interrogatives. The lowering of the weaving associated with the total contour is what caracterizes neutral interrogatives, not only its nuclear contour, which can go upwards or downwards.

In a similar line of investigation, Truckenbrodt/Sandalo/Abaurre (2009) analyze a variety of Paulista dialect (spoken in Campinas) combining production and perception data. They examine four nuclear contours in both questions and statements: statements (H+L* L%), emphatic statements (same, with higher F0), yes/no-questions (L+H* L%), and surprise questions (L*+H L%). The analysis shows that the declarative/interrogative distinction is encoded by the choice of L* vs. H* pitch accent. A distinction corresponding to English committing vs. non-committing intonation is marked by the choice of bitonal H+L vs. L+H pitch accent. Focus is signaled by the absence of a pitch accent following an early narrow focus, as well as by increased relative length of the focused constituent.

Intonation patterns of BP and EP are different in meaningful ways. Regarding yes-no questions, Frota et al. (2015) verified that southern varieties of EP show an all-rising configuration instead of the fall-rise found in the Lisbon variety; northern varieties of BP show a high tonal boundary that contrasts with the low boundary found in Rio de Janeiro and other central-southern varieties. Another difference between EP and BP comes from intonational phrasing. In BP, a low tonal boundary may mark the right edge of a prosodic domain smaller than the intonational phrase in utterances with an early focus; in EP, only the intonational phrase shows tonal boundary marking.

Considering the distribution of pitch accents in utterances, Frota et al. (2015) verify that pitch accentuation is sparse in the Lisbon variety of EP, rich in the Rio de Janeiro variety of BP. Tonal events may also differ: southern varieties of BP tend to show a denser distribution of tonal events; in the standard (Lisbon) variety of EP, the distribution is sparser than in other EP varieties.

6 Conclusion

The several sociolinguistic studies on language variation in BP reviewed in this chapter have shown how the sound system is realized. They revealed the distribution of different phonological variables in the vast Brazilian territory, allowing one to determine areas where variants occur. These studies also show that the progress of variable processes and sound change is slow in BP and it is unlikely to cause changes to the language system per se.

Variationist studies (according to Labov 1972) delineate the profile of Brazilian speakers, showing the correlation between variables under analysis and their social and structural restrictions. The studies also inform which are the most productive variants in different BP dialects.

National mapping carried out on dialectological lines (Cardoso et al. 2014) confirms the results of variationist studies on variable phonological processes in BP. The geographical distribution of variants of specific segments on such maps converges in many ways with the stratification of variants in specific speech communities.

Intonational patterns in BP have only recently received scholarly attention Yet it is possible to say that the neutral declarative pattern in Brazilian Portuguese is the same as in most known languages, and that average pitches are higher for interrogative than for declarative utterances.

7 References

Aguilera, Vanderci de Andrade (2011), *Dois momentos do /r/ retroflexo em Lavras MG: no Atlas linguístico de Minas Gerais e nos dados do projeto do Atlas linguístico do Brasil*, Diadorim 8, 125–142.

Almeida, Marco Antônio Bomfoco de (2000), *A variação das oclusivas dentais na comunidade bilíngue de Flores da Cunha: uma análise quantitativa*, Master's thesis, Porto Alegre, Pontifícia Universidade Católica do Rio Grande do Sul.

Battisti, Elisa (1997), *A nasalização no português brasileiro e a redução dos ditongos nasais átonos: uma abordagem baseada em restrições*, Doctoral thesis, Porto Alegre, Pontifícia Universidade Católica do Rio Grande do Sul.

Battisti, Elisa (2002), *A redução dos ditongos nasais átonos*, in: Leda Bisol/Cláudia Regina Brescancini (edd.), *Fonologia e variação. Recortes do português brasileiro*, Porto Alegre, EDIPUCRS, 183–202.

Battisti, Elisa/Dornelles Filho, Adalberto Ayjara (2015), *Análise em tempo real da palatalização de /t/ e /d/ no português falado em uma comunidade ítalo-brasileira*, Revista da ABRALIN 14/1, 219–244.

Battisti, Elisa, et al. (2007), *Palatalização das oclusivas alveolares e a rede social dos informantes*, Revista virtual de estudos da linguagem—REVEL 5/9, http://www.revel.inf.br, last accessed 18.02.2022.

Bernardo da Silva, Joelma C. (2011), *Descrição entoacional da questão total nos falares das capitais brasileiras*, Anais do colóquio brasileiro de prosódia da fala 1, 44–48.

Bisol, Leda (1981), *Harmonização vocálica: uma regra variável*, Doctoral thesis, Rio de Janeiro, Universidade Federal do Rio de Janeiro.
Bisol, Leda (1985), *A palatalização e sua restrição variável*, Relatório de pesquisa mimeo, Universidade Federal do Rio Grande do Sul.
Bisol, Leda (1992), *O acento. Duas Alternativas de Análise*, Manuscript.
Bisol, Leda (1994), *O acento e o pé binário*, Letras de Hoje 29, 25–36.
Bisol, Leda (1998), *A nasalidade, um velho tema*, D.E.L.T.A. 14, special issue, 1–15.
Brescancini, Cláudia Regina (2002), *A fricativa palato-alveolar e sua complexidade: uma regra variável*, Doctoral thesis, Porto Alegre, Pontifícia Universidade Católica do Rio Grande do Sul.
Cagliari, Luiz Carlos (1977), *An experimental study of nasality with particular reference to Brazilian Portuguese*, Doctoral thesis, Edinburgh, University of Edinburgh.
Callou, Dinah/Leite, Yonne/Coutinho, Lilian (2001), *Elevação e abaixamento das vogais pretônicas no dialeto do Rio de Janeiro*, Organon 5, 71–78.
Callou, Dinah/Moraes, João Antônio/Leite, Yonne (1994), *A variação de /s,r/ em posição final de sílaba e os dialetos brasileiros*, Manuscript.
Câmara Jr., Joaquim Mattoso (1970), *Estrutura da língua portuguesa*, Petrópolis, Vozes.
Cardeira, Esperança (2006), *O Essencial sobre a História do Português*, Lisboa, Editorial Caminho.
Cardoso, Suzana Alice Marcelino da Silva, et al. (2014), *Atlas linguístico do Brasil*, Londrina, EDUEL.
Collischonn, Gisela (2010), *O acento em português*, in: Leda Bisol (ed.), *Introdução a estudos de fonologia do português brasileiro*, Porto Alegre, EDIPUCRS, 132–165.
Dutra, Eduardo de Oliveira (2007), *A palatalização das oclusivas dentais /t/ e /d/ no município de Chuí*, Master's thesis, Porto Alegre, Pontifícia Universidade Católica do Rio Grande do Sul.
Espiga, Jorge (2003), *Alofonia de /l/ no sul do Rio Grande do Sul: aspectos fonéticos e fonológicos*, in: Dermeval da Hora/Gisela Collischonn (edd.), *Teoria linguística: fonologia e outros temas, João Pessoa*, Editora Universitária, 251–290.
Freire, Josenildo Barbosa (2016), *Variação, estilo, atitude e percepção linguística: o caso das laterais / ʎ/ e /l/ no falar paraibano*, Doctoral thesis, João Pessoa, Universidade Federal da Paraíba.
Frota, Sónia/Cruz, Marisa/Svartman, Flaviane/Collischonn, Gisela/Fonseca, Aline/Serra, Carolina/Oliveira, Pedro/Vigário, Marina (2015), *Intonational variation in Portuguese: European and Brazilian varieties*, in: Sónia Frota/Pilar Prieto (edd.), *Intonation in Romance*, Oxford, Oxford University Press, 235–283.
Girelli, Carl Anthony (1988), *Brazilian Portuguese syllable structure*, Doctoral thesis, Storrs (CT), University of Connecticut.
Goldsmith, John (1976), *Autosegmental phonology*, Doctoral thesis, Cambridge (MA), MIT Press.
Goldsmith, John (1995) (ed.), *The Handbook of Phonological Theory*, Oxford, Basil Blackwell.
Guy, Gregory R. (1981), *Linguistic variation in Brazilian Portuguese: aspects of the phonology, syntax, and language history*, Doctoral thesis, Philadelphia (PA), University of Pennsylvania.
Hora, Dermeval da (1990), *A palatalização das oclusivas coronais: variação e representação não-linear*, Doctoral thesis, Porto Alegre, Pontifícia Universidade Católica do Rio Grande do Sul.
Hora, Dermeval da (2003), *Reanálise da consoante em final de palavra: coda ou ataque de núcleo vazio?*, in: Claudia Roncarati/Jussara Abraçado (edd.), *Português Brasileiro II: contato linguístico, heterogeneidade e história*, Rio de Janeiro, EDUFF, 79–92.
Hora, Dermeval da (2006), *Vocalização da lateral /l/: correlação entre restrições sociais e estruturais*, Scripta 9, 31–46.
International Phonetic Alphabet (2005), *International Phonetic Association*, http://internationalphoneticassociation.org/, last accessed 18.02.2022.
Jun, Sun-Ah (2005), *Prosodic Typology: The Phonology of Intonation and Phrasing*, New York, Oxford University Press.

Kiparsky, Paul (1982), *Lexical Phonology and Morphology*, in: The Linguistic Society of Korea (ed.), *Linguistics in the morning calm*, Seoul, Hanshin Publishing Co., 3–91.
Labov, William (1966), *The social stratification of English in New York City*, Washington, D.C., Center for Applied Linguistics.
Labov, William (1972), *Sociolinguistic patterns*, Philadelphia, University of Pennsylvania Press.
Ladd, D. Robert (1996), *Intonational Phonology*, Cambridge, Cambridge University Press.
Ladd, D. Robert (^{2}2008), *Intonational Phonology*, Cambridge, Cambridge University Press.
Lopez, Barbara S. (1979), *The sound pattern of Brazilian Portuguese*, Doctoral thesis, Oakland (CA), University of California.
Magalhães, José Olímpio de (1990), *Une étude de certains processus de la phonologie portugaise dans le cadre de la Théorie du Charme et du Gouvernement*, Doctoral thesis, Montréal, Université de Montréal.
Magalhães, José Sueli de (2004), *O plano multidimensional do acento na Teoria da Otimidade*, Doctoral thesis, Porto Alegre, Pontifícia Universidade Católica do Rio Grande do Sul.
Maia, Vera Lúcia M. (1986), *Vogais pretônicas médias na fala de Natal*, Estudos Linguísticos e Literários 5, 35–47.
Mateus, Maria Helena Mira/Andrade, Ernesto (2000), *The Phonology of Portuguese*, Oxford, Oxford University Press.
Mattos e Silva, Rosa Virgínia (1996), *Sobre desencontros e reencontros. Filologia e Lingüística no Brasil no século XX*, in: *XI Encontro Nacional da ANPOLL*, unpublished.
Mohanan, Karuvannur Puthanveettil (1986), *The Theory of Lexical Phonology*, Dordrecht, Riedel.
Monaretto, Valéria Neto de Oliveira (1992), *A vibrante: representação e análise sociolinguística*, Master's thesis, Porto Alegre, Universidade Federal do Rio Grande do Sul.
Moraes, João Antônio de (1993), *A entoação modal brasileira: fonética e fonologia*, Cadernos de Estudos Linguísticos 25, 101–111.
Moraes, João Antônio de (1998), *Intonation in Brazilian Portuguese*, in: Daniel Hirst/Albert Di Cristo (edd.), *Intonation systems: a survey of twenty languages*, Cambridge, Cambridge University Press, 179–194.
Moraes, João Antônio de (2008), *The Pitch Accents in Brazilian Portuguese: analysis by synthesis,* in: Plinio A. Barbosa/Sandra Madureira/César Reis (edd.), *Proceedings of Speech Prosody 2008*, Campinas, UNICAMP, 389–397.
Moraes, João Antônio de/Wetzels, Willem Leo Marie (1992), *Sobre a duração dos segmentos nasais e nasalizados em português—um exercício de fonologia experimental*, Cadernos de estudos linguísticos 23, 153–166.
Nascentes, Antenor (1922), *O linguajar carioca*, Rio de Janeiro, Organizações Simões.
Oliveira, Marilúcia Barros de/Lima, Alcides Fernandes/Razky, Abdelhak (2009), *Imagens preliminares da realização variável de /l/ prevocálico no Estado do Pará*, Signum: Estudos Linguísticos 12/1, 257–278.
Oliveira, Marilúcia Barros de/Lima, Alcides Fernandes/Razky, Abdelhak (2016), *Palatalização de L diante de I no Português Brasileiro*, Revista Lingüística 32/2, 63–72.
Pagotto, Emílio Gozze (2004), *Variação e Identidade*, Maceió, EDUFAL.
Parkinson, Stephen (1983), *Portuguese nasal vowels as phonological diphthongs*, Lingua 61, 157–177.
Pereira, Regina C. Mendes (1997), *As vogais médias pretônicas na fala dos pessoenses urbanos*, Master's thesis, João Pessoa, Universidade Federal da Paraíba.
Pierrehumbert, Janet B. (1980), *The phonology and phonetics of English intonation*, Doctoral thesis, Cambridge (MA), Massachusetts Institute of Technology.
Pierrehumbert, Janet B./Beckman, Mary E. (1988), *Japanese Tone Structure*. Cambridge, Mass., The Massachusetts Institute of Technology Press.

Quednau, Laura (1993), *A lateral pós-vocálica no português gaúcho: análise variacionista e representação não-linear*, Master's thesis, Porto Alegre, Universidade Federal do Rio Grande do Sul.

Rosignoli, Carolina Carbonari/Fernandes-Svartman, Flaviane Romani (2016), *O padrão entoacional das sentenças interrogativas do português brasileiro em fala manipulada*, Estudos Linguísticos 45/1, 60–72.

Schwindt, Luiz Carlos da Silva/Bopp da Silva, Thaís (2010), *Panorama da redução da nasalidade em ditongos nasais átonos no português do sul do Brasil*, in: Leda Bisol/Gisela Collischonn (edd.), *Português do sul do Brasil*. Variação fonológica, Porto Alegre, EDIPUCRS, 15–30.

Seara, Izabel C. (2000), *Estudo acústico-perceptual da nasalidade das vogais do português brasileiro*, Doctoral thesis, Florianópolis, Universidade Federal de Santa Catarina.

Selkirk, Elisabeth (1982), *The syllable*, in: Harry van der Hulst/Norval Smith (edd.), *The structure of phonological representations*, vol. 2, Dordrecht, Foris, 337–383.

Silva, Ailma do Nascimento (2009), *As pretônicas no falar teresinense*, Doctoral thesis, Porto Alegre, Pontifícia Universidade Católica do Rio Grande do Sul.

Silva, Myrian Barbosa da (1989), *As pretônicas no falar baiano: a variedade culta de Salvador*, Doctoral thesis, Rio de Janeiro, Universidade Federal do Rio de Janeiro.

Silvestre, Aline P. dos S./Cunha, Cláudia de S. (2013), *Pelos cantos do Brasil. A variação entoacional da asserção neutra em Natal*, Letrônica 6/1, 179–195.

Tasca, Maria (1999), *A lateral em coda silábica no sul do Brasil*, Doctoral thesis, Porto Alegre, Pontifícia Universidade Católica do Rio Grande do Sul.

Tasca, Maria (2002), *A interferência da língua falada na escrita das séries iniciais*, Porto Alegre, EDIPUCRS.

Teyssier, Paul (1980), *História da língua portuguesa*, São Paulo, Martins Fontes.

Truckenbrodt, Hubert/Sandalo, Filomena/Abaurre, Maria Bernadete M. (2009), *Elements of Brazilian Portuguese intonation*, Journal of Portuguese Linguistics 8/1, 77–115.

Verney, Luís António (1746),*Verdadeiro método de estudar*, Valensa, Oficina de Antonio Baile.

Viegas, Maria do Carmo (2001), *O alçamento das vogais médias pretônicas e os itens lexicais*, Doctoral thesis, Belo Horizonte, Universidade Federal de Minas Gerais.

Vieira, Maria José Blaskovski (1994), *A neutralização das vogais pós-tônicas*, Master's thesis, Porto Alegre, Universidade Federal do Rio Grande do Sul.

Vieira, Maria José Blaskovski (2002), *As vogais médias postônicas: uma análise variacionista*, in: Leda Bisol/Cláudia Regina Brescancini (edd.), *Fonologia e variação. Recortes do português brasileiro*, Porto Alegre, EDIPUCRS, 127–160.

Vigário, Marina (2003), *The prosodic word in European Portuguese: Interface Explorations*, Berlin/New York, Mouton de Gruyter.

Votre, Sebastião J. (1978), *Variação fonológica no Rio de Janeiro*, Doctoral thesis, Porto Alegre, Pontifícia Universidade Católica do Rio de Janeiro.

Weinreich, Uriel/Labov, William/Herzog, Marvin (1968), *Empirical foundations for a theory of language change*, in: Winfred Philipp Lehmann/Yakov Malkiel (edd.), *Directions for historical linguistics*, Austin, University of Texas, 95–195.

Wetzels, W. Leo M. (1997), *The lexical representation of nasality in Brazilian Portuguese*, Probus 9/2, 203–232.

Wetzels, W. Leo M. (2011), *The Representation of Vowel Height and Vowel Height Neutralization in Brazilian Portuguese*, in: John Goldsmith/Elizabeth Hume/W. Leo M. Wetzels (edd.), *Tones and Features*, Berlin/Boston, De Gruyter, 331–359.

Williams, Edwin B. (1968), *From Latin to Portuguese: Historical Phonology and Morphology of the Portuguese Language*, Philadelphia, University of Pennsylvania Press.

Carlos Alexandre Gonçalves

13 Morphology

Abstract: This chapter outlines the distinction between lexical and inflectional morphology in contemporary Brazilian Portuguese (BP) through a description of synchronic paradigms. BP shares many morphological features with other varieties, such as European Portuguese (EP) and Portuguese spoken in Africa and Asia, but it also displays properties that set it apart from other members of the Lusophone world. In this chapter, we privilege the latter aspects, as well as discussing new syncretisms and morphological reductions in the contemporary spoken language in comparison to written language. In particular, emphasis is given to the boundaries between compounding and derivation. In mapping differences between these two processes, we describe several cases that could be interpreted as derivation or as compounding: prefixing, adverbs formation in *-mente* (*-ly*, in English) and diminutives, among others. In addition, we discuss some types of word-formation (henceforth WF) not addressed in the grammatical tradition, such as blending, clipping and other non-concatenative morphological processes, especially recent patterns of reduplication and acronyms used today in BP. While influential traditional studies, such as Câmara Jr. (1970), Basilio (1987) and Sandmann (1989), are discussed where appropriate, the description in this chapter is mainly based on recent works of BP morphology and intends to show the arena of current morphological discussions.

Keywords: inflection, derivation, composition, diminutives, word formation, composition-derivation-continuum, compounding

1 Introduction

In this chapter, we intend to describe some specificities of Brazilian Portuguese in relation to other varieties of Portuguese with regard to issues of synchronic morphology. We will focus on aspects of Brazilian Portuguese linked to inflection, word formation, the creation of diminutives etc., always intending to make flexible the borders between the main areas of morphology: inflection, derivation and compounding.

The description is structured as follows: Section 1 explores differences between inflection and derivation, and addresses the derivational properties of grammatical categories such as gender in nouns, and tense-mood-aspect in verbs. Section 2 turns to the formation of diminutives, one of the most controversial cases of categorization in BP morphology, after which the boundaries between compounding and derivation are described and the typological diversity of the BP WF system is addressed. The study then moves on to processes such as lexical blending, neoclassical compound-

https://doi.org/10.1515/9783110405958-014

ing, re-compounding, and the creation of new morphemes (splinters). Finally, the diversity of nonconcatenative processes of BP is discussed.

2 Inflection and derivation

Inflection and derivation can be understood as two branches of morphology. The first regards the part of morphology that is relevant to syntax; thus, inflection is concerned with syntactically-driven morphology. Conversely, derivation arises from the need to create new meanings, but it is also constrained by syntax in some way, and hence is more than merely that part of morphology used to create new lexical items.

Most work in this area has focused on the fact that the difference between inflection and derivation is a discrete one, highlighting formal criteria to make the distinction. Some of these formal criteria are listed in Table 1, based on Gonçalves (2005; 2011), who noted almost thirty differences between these processes. Despite various properties that undoubtedly characterize the two "morphologies", clear-cut distinctions concerning inflection and derivation have not been found, which points towards the existence of a scale between them.[1]

Table 1: Summary of the differences between inflection and derivation.

Inflection	Derivation
A. Obligatory	A. Optional
B. Conditioned by the syntax	B. Not conditioned by syntax
C. Highly structured	C. Idiosyncratic
D. Closed class	D. Open class
E. Meanings are manifested only morphologically	E. Meanings are manifested by means of various linguistic forms
F. There are fewer inflectional morphemes in a language than derivational ones	F. There are more derivational formatives in a language than inflectional ones
G. Does not change the stress of the source form	G. Often changes the stress of the source form

1 In this table, properties such as those described in G, K and U, for example, could easily be questioned by someone who works with typology, using counterexamples of Russian, Japanese and Bantu. In Russian, the accent (criterion in G) changes very often along its flexion; speaker's characteristics (age, sexual orientation, region and education) is widely represented in morphological and morphosyntactic paradigms in Japanese (K); in many Bantu languages, flexion occurs on the left edge and not on the right (M). These differences, therefore, should be viewed as general trends (not as a final verdict on the status of morphological units).

Table 1: (continued)

Inflection	Derivation
H. Does not express the speaker's perspective about something or someone	H. Can express the speaker's perspective about something or someone
I. Produces various forms of the same word	I. Creates new lexical units
J. Word class does not change	J. Sometimes changes word class
K. Does not identify the speaker by age, sexual orientation, region or education	K. Sometimes identifies the speaker by age, sexual orientation, region or education
L. An inflectional rule cannot be reapplied	L. A derivational rule can be reapplied
M. Expression at word periphery	M. Expression close to the base
N. Presents constant meanings	N. Presents variable meanings
O. Usually no gaps in the paradigm, being fully productive in a word class	O. Usually has a gap in the paradigm. Some derivational processes can be highly productive, but many are restricted to a part of the vocabulary
P. Fusion cases are more common in meanings linked to inflection	P. Fusion cases are less common in meanings linked to derivation
Q. Never changes the reference because its meanings are less relevant to the meaningof the base	Q. Is more relevant to the meaning of the base andalways changes the reference
R. The absence of inflectional marks (suffixes in BP)produces incomplete words	R. The absence of derivational affixes often produces complete words (roots orthemes)
S. Grammatical meanings can co-occur in the same morph; therefore, cumulativeexpression is possible in inflection	S. Always non-cumulative
T. Is less sensitive to lexicalization phenomena	T. The loss of compositionality is common (more sensitive to lexicalization)
U. The head is on the left (pattern nucleus-modifier)	U. The head is on the right (pattern modifier-nucleus)

It is usually claimed that inflection is used to create grammatical forms of the same word (I), without changing its word class (J), while derivation necessarily generates new lexical units (I), changing the reference, that is, pointing to a new designatum (Q). However, Table 1 raises the following questions:

(a) When does a new word arise and in what conditions is the same word only modified grammatically? What makes us believe that, for instance, *perseguir* 'to pursue' and *perseguido* 'pursued'[2] are different words, and that *perseguida* 'pursued, fem.' is only a gender inflection of *perseguido* 'pursued, masc.'?
(b) What criterion (or criteria) can be used to differentiate a new word from the same word modified for gender, number or degree? For instance, are *janela* 'window', *janelinha* 'little window' and *janelão* 'big window' new words or variations of the same word?

The distinction between what is or is not a new lexical item is not always clear in practice. In BP, for example, *perseguida* 'pursued, fem.' acquires an independent meaning and can refer pejoratively to the female genitals. The nominal forms of verbs (gerund, participle past and infinitive) do not have a stable categorization. The affixes responsible for the categorization of these forms create units with unstable behavior in its word class. The following examples show the instability of participle forms:

(1) O homem foi *roubado*. (verb)
'The man was robbed'

A bolsa *roubada* foi encontrada. (adjective)
'The stolen bag was found'

Ele está numa *roubada*. (noun)
'He is in a fix'

Some evidence for this instability is that there is no unanimous agreement as to what is to be treated as a new word or as a form of the same word. Sandmann (1989, 31) shows that the classification of a word as a standalone unit or a formal variation of a word is sometimes difficult even for lexicographers. In this regard, Basilio (1987, 12) raises the question: 'should we consider *perdido* 'lost' as a verbal form of *perder* 'to lose' or as another word?'[3] Sandmann (1989, 31) notes in this sense some contradictions in the most influential of Portuguese dictionaries, the *Aurélio*:

> 'The Aurélio records the following past particles as distinct from the verb entries: *folgado* ('rascal'), *apaixonado* ('in love'), *cansado* ('tired'), *lido* ('read'). Immediately after the entry in brackets for the verbs *folgar* ('to hang out'), *apaixonar* ('to fall in love'), *cansar* ('to tire') and *ler* ('to read'), it sorts the corresponding past participles as adjectives. Aurélio does not cite as adjectives *come-*

2 As in English, the past participle may be used as a verbal form or as an adjective. Even if the word is used in the verbal form, it will vary in gender, marked by *-o* or *-a* (as occurs in the passive voice).
3 Original: Poderíamos considerar perdido como forma verbal de 'perder' ou como outra palavra?

çado ('started') and *chateado* ('upset'), for example, which is particularly puzzling in the case of *chateado*, certainly used far more as an adjective than as a past participle.'[4]

Gonçalves (2005) shows that in BP there are always counterexamples in terms of the criteria used to argue for a discrete opposition between inflection and derivation. Those who propose a dichotomy between the two "morphologies" give different weight to each criterion. If all the criteria are analyzed with one measure and the data are effectively observed, one can imagine a cline between inflection and derivation. Altakhaineh (2014) observes that to argue for the split morphology, researchers generally use intuition as their basis—which, according to him, is not an argument—and they "close their eyes" to counterexamples. In what follows some of the questions raised by Altakhaineh (2014) will be shown through BP data.

As examples of failures in the criteria used to distinguish inflection from derivation, Altakhaineh (2014) points to syntactic relevance (differences in A and B). As in inflection, the agreement leading to the use of derivational affixes, such as those listed in Table 2, is evidence of their syntactic relevance: word-class switching, in derivational processes, must also be regarded as essential to syntax. Moreover, the author questions syntactic relevance in categories like mood, tense and aspect, which are in no way imposed by syntax.

Table 2: Short list of suffixes that change word class in BP.

Verb → Noun	Verb → Adjective	Noun → Adjective	Noun, Adj. → Verb	Adjective → Adverb	Adjective → Noun
-ção; *-ada*	*-vel*; *-nte*	*-ense*; *-ar*	*-izar*; *-escer*	*-mente*	*-ice*; *-idade*
inibição 'inhibition'	*gerenciável* 'manageable'	*canadense* 'Canadian'	*agilizar* 'streamline'	*felizmente* 'fortunately'	*esquisitice* 'oddity'
esticada 'stretch'	*estafante* 'gruelling'	*hospitalar* 'hospital'	*florescer* 'flourish'	*certamente* 'certainly'	*felicidade* 'happiness'

4 The following gloss, from the Aurélio Dictionary (Ferreira, 1987), illustrates Sandmann's comments: *folgado* ('cheeky', in relation to the verb *folgar*) appears in a separate gloss of the verb *folgar* 'to hang out', but its meanings do not relate to the meaning of the verb. Moreover, it is substantive in the most common meaning (3). 'Folgado: *adjetivo* **1**. livre de tarefas, de deveres; descansado. 'depois que defendeu a tese, está bem mais folgado' **2**. largo, amplo. 'o sapato ficou folgado'. **3**. Esperto. É muito folgado. Quer se dar bem em tudo'. ('Cheeky: adjective 1. free of duties; rested. After he defended the thesis, it is much more rested 2. wide; comfortable. 'the shoe is comfortable' 3. Abused; scoundrel. He is very roguish; always fools people').

As for the obligatory nature of the expression in inflection (A), Bybee (1985) demonstrates that morphological zeros express grammatical meaning through the absence of formal expression. If the mark is ∅, it means there is no compulsory expression in inflectional morphology.

Bybee (1985) provides cross-linguistic evidence that disturbances in morph-morpheme mapping are not entirely arbitrary. For her, form-meaning mapping problems in morphology, such as ∅ morph, involve basic (or cognitively simpler) grammatical categories, and such categories often fail to present formal expression in natural languages. The absence of phonetic expression, she argues, is psychologically motivated since zeros determine non-marked categories or grammatical meaning from which others can be explained.

In Portuguese, singular form, masculine gender, present indicative tense and third person singular are grammatical meanings which are not expressed phonetically. Thus, only unmarked grammatical meanings are represented by zeros in BP: the ∅ morph is, in all these cases, the most general member of their grammatical categories[5]. For example, in 2, a masculine form (*aluno* 'student, masc.') can be used with reference either to a man or a woman; in 3, the inflected verb (*começa* 'begin') does not necessarily express the present tense. Note that an adverb can be used to indicate tense:

(2) *Aluno* é difícil de entender.
'A student is hard to understand'—man or woman

(3) *Começa* hoje o festival.
'The festival starts today'
Começa amanhã o festival.
'The festival starts tomorrow'
Começa o festival.
'Festival starts' (in a newspaper report)

Spencer (1998) notes that, when one addresses the paradigmatic organization of inflection (unlimited applicability), cases of defectiveness are forgotten (difference in C). Defectiveness is often justified in terms of the same reason that determines the incompleteness of derivation: semantic incompatibility. For example, verbs that express a natural phenomenon have no meaning compatible with morphological markers

5 As Bybee herself points out, the association of zero-morph with an unmarked element does not have definitive power, because there are languages, such as Russian, that the use of ∅ in inanimate feminine and neutral words represents the plural genitive. However, the use of zero to express the unmarked member of the pair is much more usual.

which require a subject (⋆*chovemos* 'we rain'; ⋆*nublei* 'I clouded').[6] Thus, number/person affixes present constraints on this applicability and, therefore, distance themselves from the possibility of becoming established as inflectional by the criterion of generality (O).

Bybee (1985) raise questions regarding the distinction between inflection and derivation as to aspects of meaning. Aronoff/Fudeman (2005), among others, argue that, in inflection, there is no expression of a new concept: inflected forms indicate the same concept as the base (Q). Instead, in derivation, words are formed to designate new concepts. Thus, according to the criterion of compositionality (T), in inflected forms the meaning is always transparent because the content of the word shape is indicated by the sum of its parts, while derivative forms may have non-compositional interpretations.

In BP, numerous cases exist of non-compositional meaning in inflection, both in nouns (inflected forms in gender and number), and in verbs (forms varying in tense-mood-aspect and number-person).[7]

Concerning the feminine morphological marker, the unstressed final vowel *-a* ([ɐ]), as in *professor+a* 'teacher+fem', there are in BP various semantically lexicalized morphological structures in which this vowel does not express the notion of female. The word *peru+a* 'turkey+fem', for example, does not necessarily mean 'female turkey', and can mean 'a type of vehicle' 'van' or 'tart' 'an overdressed woman'. Designations for women in colloquial speech community reveal strong use of semantic lexicalization; thus *coelh+a* 'rabbit+fem' means 'teenager who has several children'. The recent formation *preparad+a* 'prepared+fem' illustrates how women can be described negatively in the context of night clubs, in that it refers to a 'woman who goes to a club with no underwear, ready to engage in sexual relations'. In 4, other feminine forms appear lexicalized:

(4)	brux+a	'witch+fem' = 'moth'
	porc+a	'pig+fem' = 'nut for a bolt'
	caminhoneir+a	'truck driver+fem' = 'lesbian'

Lexicalization also exists involving the plural (PL) mark, a fricative /s/, although to a lesser extent:

6 In order to standardize the symbols used in this chapter with those employed in the diachronic morphology chapter, we use the symbol ⋆ to indicate ungrammaticality and reserve * for reconstructed historical data.

7 In Portuguese, the verbal markers are always cumulative (criterion in S): proximity to the root indicates simultaneously mood/tense/aspect, and distance from it expresses, at the same time, number/person.

(5) costa+s 'coast+PL' = 'back'
copa+s 'pantry+PL' = 'in cards, the suit of hearts'
féria+s 'income+PL' = 'holiday'

Vivas (2015) notes that in verbs there are numerous cases of modified forms in tense-mood-aspect (TMA) and/or number-person (NP)[8] that can act on other morphosyntactic categories also undergoing changes in meaning. In the following data, there are cases of inflected forms, which, besides being verbs, can also be categorized in other word classes:[9]

(6) Interjections: *tomara*, *dera*, *soubera*, *quisera*, *pudera* (case 1)
(These forms reveal the speaker's desire for something to happen in different ways)
Interjections: *demorou*, *formou*, *partiu*, *valeu* (case 2)
(These forms reveal the speaker's contentment about something said by someone)
Adjectives: *cheguei*, *choquei*, *caguei* (case 3)
(These forms reveal the negative impression of the speaker about something or someone)
Discourse markers: *sabe*, *entende*, *olha* (case 4)
(These forms refer to the interaction between speaker and listener)

Words modified in the MTA and NP can serve to indicate the speaker's viewpoint, contrary to the expectations in H and K. When this happens, the syntactic category of the word changes. The first two cases (interjections) involve forms with a marking of TMA that have fallen into disuse (*-ra*, inflectional affix indicating the pluperfect indicative is no longer used in spoken language) or that are inflected with an affix of number-person which does not occur in all conjugations (*-u*, the third-person singular formative appears exclusively in the simple past indicative).[10]

8 The template of Portuguese verbs includes the constituents *theme* and *inflection*, with the former comprising the root and a thematic vowel (/a/ in the first conjugation; /e/ in the second, and /i/ in the third), and the latter comprising the formal variation in the cumulative dimensions of mood-tense-aspect (MTA) and number-person (NP), in this sequence:

Theme		Inflection	
Root	Thematic Vowel	MTA	NP

9 Since these words are slang, it is difficult to assign a precise meaning to them in that their uses are pragmatic. For now, we prefer to indicate a generic meaning for each group.

10 BP has three different kinds of tenses to express the notion of past, with morphological marking. These tenses indicate the following aspects: perfective (a), imperfective (b) and pluperfective (c).
(a) *Quando Jane telefon***ou**, *John estava dormindo*. 'When Jane telephoned, John was asleep.'
(b) *Sempre que Jane telefona***va**, *John estava dormindo*. 'Whenever Jane telephoned, John was asleep.'
(c) *João já saí***ra**, *quando Jane telefonou*. 'John had already left, when Jane telephoned.'

In the third case, evaluative adjectives are formed through verbs modified in the first-person singular of the present indicative. Also, in this case, a less general formative appears in the verbal conjugation: the morph *-i*, first-person singular particle, only occurs, in indicative mood, in the present tense and in the future present. The first case interests us here, in that there is allomorphy, given that the thematic vowel /a/, the first conjugation affix, becomes [e] before [i].

In the latter case, verbal forms are equivalent to discourse markers and, as they are inflected in the third person singular, present indicative tense, are formally devoid of NP and TMA markers (both the present tense and the third person singular are ∅ in the verbal paradigm). A summary of the morphological behavior of re-categorized verbs in BP is provided in Table 3:

Table 3: Summary of the behavior of re-categorized verbs.

	TMA	NP
Case 1	*-ra*	∅
Case 2	∅	*-u*
Case 3	∅	*-i*
Case 4	∅	∅

The existence of ∅s, the use of markers that have fallen into disuse and those that are not general in verbal paradigm, tend to create forms with a greater predisposition to function in another class (interjections such as *tomara* 'I hope so!', and *valeu* 'cool!', or adjectives, like *cheguei* 'extravagant' and thus to acquire a non-verbal meaning, such as a non-verbal kind of behavior. In other words, it seems that these aspects—the use of a marked form, its disuse, or its absence—give words a less verbal appearance, which are more likely to acquire other uses and meanings in spoken language.

Bybee (2010) distinguishes compositionality from analyzability. Even considering these correlated and gradually distinct parameters, she demonstrates the need to make a differentiation here as a means of better understanding how the linguistic constructions vary. Analyzability is related to the existence of segmentation of a word into smaller units provided by phonetic form and meaning. Compositionality, in turn, consists of identifying more than one meaning in the word. This multiple content is not related to phonetic sequences and cannot be divided into discrete parts of form and meaning. Idioms and compounds are often analyzable even when not compositional

BP has also two kinds of future tenses, with morphological markers. Simple future expresses certainty, assurance (d); conditional future expresses an action unlikely to occur (e).
(d) *Eu i***rei** *à festa amanhã*. 'I will go to the party tomorrow.'
(e) *Eu i***ria** *à festa amanhã, se pudesse*. 'I would go to the party tomorrow, if I could.'

—for example, *baba-ovo*'brownnoser', a compound, can be analyzed as formed by the words *babar* 'slaver' and *ovo* 'scrotum', although its interpretation is holistic.[11]

With examples of supplementary forms in verbal inflection, we can say that the opposite is also true: there is compositionality, but not analyzability. In *foi* 'went', we know that it indicates 'past' and 'third-person' even though this form is not divided into smaller portions of form and meaning. The analyzability relates to the existence of segmentation with particular meaning in the word; the compositionality, in turn, consists of identifying more than one meaning within the word.

When verbal forms express a positive view of a situation, as in the examples below, extracted from Vivas (2015), they fail to convey the meanings of past indicative tense and third-person singular. In other words, there is a loss of compositionality and analyzability. In 7, B agrees with A and welcomes the idea of leaving tomorrow.

(7) A—Vamos sair amanhã? 'Let's go out tomorrow?'
B—*Fechou*! 'Sure!'

The same is true in *roupa cheguei* 'garb'; 'I arrived' = 'very flashy clothes', in which the meanings of first-person singular and perfect indicative tense, in *cheguei*, are neither analyzable nor compositional. The function of *cheguei* here is to show the speaker's derogatory impression of the clothes. This pattern is productive only in spoken language and serves to demonstrate a negative impression of something or someone. Another example, *verde choquei* 'green'; 'I shocked' = 'very intense green', indicates the speaker's opinion that the color tone is extravagant.

Bybee (2010) demonstrated the role of frequency in the loss of compositionality and analyzability. According to her, semantic and pragmatic changes that reduce compositionality are aided by frequency or repetition, but the origin of these changes is in the contexts where the complex unit is used. Repetition, in certain contexts, with certain forms of association with their meanings, is what would trigger semantic and pragmatic change. If a sequence of morphs or words is increasingly used, it will be stronger as a unit. Thus, there is a reduction in association with the component parts. The loss of association with these component parts causes an increase in the autonomy of the sequence. It is no coincidence that these are lexicalization patterns in verbal inflection characterized by a change in the thematic vowel of the first conjugation, the most productive one. This leads us to think of a constructional block, *-ou* (allomorphic thematic vowel + third-person singular mark) and *-ei* (allomorphic thematic vowel +

11 Bybee does not "oppose" the two notions, but treats them as "closely related" (Bybee 2010, 44). I agree with the reviewer of this study as to the analyzability of idioms and compounds, but it is questionable to talk about compositionality in the case of suppletion. Bybee (2010, 45) defines compositionality as "the degree of predictability of the meaning of the whole from the meaning of the component parts". This dichotomy, however, already appears in accounts by cognitive linguists such as Langacker (1987).

first-person singular mark), the most typical in spoken language. The sequence is integrated in such a way that it went on to form a unit, i.e., a chunk.

Interjections and adjectives in 6 reveal the pragmatic impact of the speaker that emits a value judgment about the utterance. Interjections, such as *formou* 'sure!' and *arrasou* 'wonderful!', evaluate an event positively while the adjectives *cheguei* and *choquei* indicate negative judgment on the part of the speaker. Thus, the idea that the inflection is completely transparent (criterion in N) and immune to changes of meaning (I) or expressive effects (H) and the creation of lexicalization patterns (T) is not true, at least for BP.

For the aforementioned reasons, the criteria listed in Table 1 do not behave consistently and accurately. Therefore, the mapping of those aspects that differentiate inflection from derivation should be seen as an attempt to diagnose affixes and not as a verdict on their morphological status. BP data show that the difference between the two morphologies, if in fact it exists, is a matter of degree only. In other words, inflection and derivation do not involve a discrete opposition, but a gradient one, and they can be interpreted as morphological processes whose boundaries are not always clear. In Cognitive Linguistics (Langacker 1987), the hypothesis that there is no rigid separation between lexicon and grammar reinforces the scalar nature from inflection (grammar) to derivation (lexicon). The idea discussed here is compatible with the notion of prototypes, originally developed by Rosch (1978), since there are more central and more peripheral members in each of these processes. These two "morphologies" can be interpreted as prototypical poles on a scale, and for this reason there are several morphological elements that do not fit well on either side.[12] In the following section, a difficult case to categorize in BP is discussed: X-*(z)inho* diminutives.

3 A problematic case of categorization: diminutives

Evaluative affixation is one of the most productive domains in BP word formation (WF). Since it is a resource used primarily in spoken language, it is studied only at quite a superficial level in schools. In this section, we describe the most interesting case of evaluative suffixation: the creating of diminutive forms, especially those ending in *-inho*/*-zinho*.

The status of *-inho* and *-zinho* is very controversial in the literature. The proposals of categorization here go from inflection to compounding, through to derivation. For example, Pereira (1908) considers the formation of diminutives to be an inflectional process due to a historical issue related to classification in Latin. Moreover, Câmara Jr. (1970) interprets these forms as derivational. He seeks to prove their derivational sta-

12 Other approaches to inflection and derivation in BP can be found, among others, in the following descriptions: Rocha (1994), Rosa (2000) and Colnaghi (2006).

tus of degree based on only two criteria listed in Table 1: (B) Not conditioned by the syntax, and (C) Idiosyncratic. In fact, for these two reasons, these formatives are categorized as derivational units.

Other authors, such as Lee (1997), do not consider *-inho* and *-zinho* to be allomorphs of the same morpheme, and interpret *-inho* as an affix (unit of derivation) and *-zinho* as a word (unit of compounding). Some authors adopt a non-discrete categorization approach. Thus, Gonçalves (2005) considers that the elements described herein are on the border between inflection and derivation, whereas Andrade (2013) analyzes them as border units between derivation and compounding. Finally, authors such as Villalva (2000) consider that such elements are actually independent units of grammar, requiring an analysis distinct from that of traditional WF processes.

This chapter will not discuss the status of diminutive forms; rather, we intend to show the morphological features that make this process so special in BP, and we will also therefore discuss the question of why they have such divergent interpretations.

There are three prominent responses to these issues. The first concerns the wide range of meanings that these elements offer. The second, that they present unique phonological properties that no other Portuguese suffix presents. The final response refers to the fact that these elements are not lexical heads like others suffixes (see criterion U, Table 1): they are neither syntactic heads (do not change the word class), morphological heads (do not assign gender), nor semantic heads (are not core meaning). Thus, as already noted above, these elements encourage different interpretations because they have characteristics that make them unique in comparison to most prototypical elements of each category with which we try to link them. Let us look at these properties.

The diminutive markers (henceforth DIM) change the base to which they are adjoined according to a range of semantic features related to different value judgments, so that their true meaning depends on pragmatic circumstances. DIM have more central meanings, which may extend to less central meanings, and also, because of this, they end up being compatible with so many different word classes.

According to Silva (2006), at the semantic centre of DIM are two dimensions to be assigned to *-(z)inho*. The first concerns the most referential meanings, which are those considered to be decreasing (diminishes) and reinforcing ones (explainers). The second dimensions concern the metaphorical and metonymic extensions that these diminutives can create:

Table 4: Central meanings of diminutives.

Central meanings		Metaphorical and metonymic extension		
diminishes	explains	size (space) to time	time to intensity	intensity to quantity
decreases the size of the *designatum*	reinforces the size of the *designatum*	of very short duration	low intensity; weak	small amount of; little
pezinho	*migalhinha*	*visitinha*	*chuvinha*	*bocadinho*
'little foot'	'small crumb'	'brief visit'	'light rain'	'very small bit'

From these most central meanings, specializations emerge from the DIM. These are divided into different levels of lexicalization depending on the speaker's beliefs and attitudes.

In sum, if the basic use indicates the scant or negligible weight of something, the pragmatic uses reveal the speaker's perspective, i.e., the speaker gives their explicit opinion about something or someone. The word *lembrancinha* 'low-value gift', for example, shows the speaker's posture, revealing that this gift serves solely as a means of not forgetting a commemorative event. Furthermore, Brazilian culture specifically values modesty as a face-saving strategy: negative politeness, in Brown/Levinson's terms (1983).

So, on the discursive level, DIMs are characterized also by professional uses. Often employed in advertising or product sales, these diminutives illustrate the reader/listener's beliefs in the quality of the product so that s/he can then proceed to purchase it, as shown in the example *Esse é o reloginho que vem junto com essa Melissinha* 'this watch comes with the sandal'. In this case, DIM smooth the referents watch+DIM *relógio* and shoes (designated by the brand 'Melissa+DIM' metonymically). Other examples of pragmatic and discursive uses of DIM are listed in Table 5:

Table 5: Pragmatics of diminutives.

Affective uses	Hedonic uses	Positive evaluations	Negative evaluations	Intensive uses
affective and compassionate	what we like	positive impression	negative impression	Increase or maximum
mãezinha	*vinhozinho*	*corpinho*	*timinho*	*cheiinho*
'dear mother'	'good wine'	'sculptural body'	'bad team'	'full to the brim'

DIM take on other semantic values, these always linked to the central idea 'small', or simply having an independent function or designating something that goes beyond its basic meaning. However, it almost always involves formations with the status of inde-

pendent lexical items purchased individually, assigning new entities in a diminutive relation to the word base. Parallel to formations in which the relationship between the derivative and the base is of a reduced size 'small type of', as in 8a, others are in a different ratio, but usually also related to the idea of 'small': similarity relations or imitation (8b), contiguity (8c) and type (8d):

(8) a. tesourinha 'small scissors'; 'scissors used to cut nails'
b. camisinha 'small T-shirt'; 'condom'
c. bandeirinha 'small flag'; 'linesman'
d. barzinho 'small bar'; 'pub'

Because of the large semantic variability here, DIM applies almost without restriction: these suffixes can adjoin to all kinds of bases, either adjectival (9a), nominal (9b), adverbial (9c), or even to interjections (9d), numerals (9e) and pronouns (9f). Henceforth, EVAL indicates a speaker's evaluation with the use of DIM:

(9) a. fininho 'fine+EVAL'
b. livrinho 'book+EVAL'
c. cedinho 'early+EVAL'
d. adeusinho 'goodbye+EVAL'
e. cenzinho 'one hundred+EVAL'
f. euzinho 'I+EVAL'

In what follows we analyze the phonological and morphological characteristics of DIM that make them so unique compared to other suffixes. In BP, *-inho* and *-zinho* present the following distribution: vowel-initial suffix (*-inho*) is adjoined to roots (*vid+a* → *vidinha* 'life', 'life+EVAL'); [z]-initial suffix (*-zinho*) is adjoined to words (*sol* → *solzinho* 'sun', 'sun+EVAL'). The variation of these two competing forms is dialect-specific and is also prone to depend on the speaker's preference, although some grammatical constraints also apply.

The most obvious constraint is the impossibility of adjoining *-inho* to athematic bases (roots that do not end in an unstressed vowel ([ɐ, I, ʊ]), like *sofá* 'sofa', *colar* 'necklace' and *pão* 'bread'. Athematic words only allow *-zinho* suffixation, displaying in the suffix the unmarked thematic vowel (*-o* [ʊ], for the masculine, and *-a* [ɐ] for the feminine). In these cases, there is agreement with the gender as to the base (masculine, in 10a, and feminine, in 10b). Bisol (2000) generalizes this property of DIM, arguing that the conditions to avoid onsetless syllables govern the choice between these two forms. Thus, she analyzes /z/ as an epenthetic consonant.

(10) a. caju$_{masc}$+zinh+o$_{masc\ unmarked\ vowel}$ 'cashew+EVAL'
b. orfã$_{fem}$+zinh+a$_{masc\ unmarked\ vowel}$ 'orphan+EVAL'

The two forms of DIM preserve the phonetic and phonological features of the base. First, they block the mid-vowel neutralization[13] for preserving lower-mid vowels ([ɛ, ɔ]), as we seen in 11c and 11f. With other suffixes, this rule always applies (11a, 11b, 11d, 11e). Note that lower-mid vowels of the base alternate with their corresponding upper-mid vowels when they move to a pre-stress syllable (in phonetic transcriptions, we use the pronunciation of the city of Rio de Janeiro):

(11) a. vela ['vɛ.lɐ] 'candle' → velório [ve.'lɔ.ɾjʊ] 'funeral'
b. vela ['vɛ.lɐ] 'candle' → velar [ve.'lah] 'to wake'
c. vela ['vɛ.lɐ] 'candle' → velhinha [vɛ.'lĩ.ɲɐ] 'candle+EVAL'
d. porta ['pɔx.tɐ] 'door' → porteiro [pox.'te.ɾʊ] 'porter'
e. porta ['pɔx.tɐ] 'door' → portaria [pox.'ta.ɾi.ɐ] 'lobby'
f. porta ['pɔx.tɐ] 'door' → portinha [pɔx.'tĩ.ɲɐ] 'door+EVAL'

DIM also preserves the phonetic nasality of stressed vowel (12a) while other related morphologically forms that also promotes stress change never do so (12b, 12c).

(12) a. trama ['tɾẽ.mɐ] 'plot' → traminha [tɾẽ.'mĩ.ɲɐ] 'trama+EVAL'
b. trama ['tɾẽ.mɐ] 'plot' → tramado [tɾa'ma.do] 'plotted'
c. trama ['tɾẽ.mɐ] 'plot' → tramarei [tɾama.'ɾej] 'I will machinate'

Like Bisol (2010), we also consider *-inho* and *-zinho* variants in complementary distribution. The presence of /z/, an epenthetic segment that appears in other WF processes, as we see in 13, serves to satisfy structural requirements.

(13) **Augmentatives** (*-ão ~ -zão*): cachorr***ão*** 'big dog', trator***zão*** 'big tractor'
Locatives (*-al ~ -zal*): mangueir***al*** 'mango plantation', cafe***zal*** 'coffee plantation'
Collectives (*-ada ~ -zada*): boi***ada*** 'a herd of cattle', guri***zada*** 'group of youngsters'
Tree names (*-eiro ~ -zeiro*): coqu***eiro*** 'coconut tree', inga***zeiro*** 'inga tree'

13 In BP, there is contrast between the mid-vowels ([e, ɛ], [o, ɔ]) only in stressed syllables: *boto* ['bo.tʊ] 'Amazon dolphin', and *boto* ['bɔ.tʊ] 'I put'; *pelo* ['pe.lʊ] 'fell', and *pelo* ['pɛ.lʊ] 'I scratch'. In pre-stressed syllables, the contrast is lost and the most common pronunciation, in the southern dialects, is the upper-mid vowels, [e, o]: *pretende* [pɾe'tẽʲdʒɪ] 'He intends'; *procura* [pɾo.'ku.ɾɐ] 'He searches'.

In the scope of WF, BP tends to prevent the creation of hiatus and to undo it when it might occur.[14] Also, DIM tend to avoid the formation of hiatus, favoring (i) the deletion of the thematic vowel—*feir+a* → *feirinha* 'market +EVAL'—or (ii) the insertion of /z/ when the base has no thematic vowel, as in *irmã* → *irmãzinha* 'sister +EVAL'.

To ensure greater fidelity to the base, DIM do not allow resyllabification, as in other derivational processes (*papel* [pa.'pɛʊ̯] → *papelaria* [pa.pe.la.'ri.ɐ] 'stationary store'), i.e. DIM preserve the structural position of the segments of the base to which they are adjoined, so that the segment in the coda in the base remains in this position in the complex word, as in *papel* [pa.'pɛʊ̯] → *papelzinho* [pa.pɛʊ̯.'zĩ.ɲʊ] 'paper+EVAL'.

Thematic words can present the epenthetic consonant /z/ when they end in *-inh*, which is phonetically identical to the suffix. Such is the case with *vinho* 'wine', whose diminutive always presents /z/: *vinhozinho* 'wine+EVAL'.

Also in relation to the principles that govern linguistic structures, DIM avoid redundant information in the grammatical gender, so that *-inho* is privileged when the thematic vowel carries information about the gender, which is deleted, as in *garot+o* → *garotinho* 'boy+EVAL', while *-zinho* is favored when there is disparity between the thematic vowel and the gender (14a) or when there is lack of gender information (14b):

(14) a. masculines end in *-a*: ***o*** *arom**a*** → **o** *aromazinh**o*** 'aroma+EVAL' feminines end in *-o*: ***a*** *trib**o*** → ***a*** *tribozinh**a*** 'tribe+EVAL'
b. genderless names: ***o***/***a*** *pianista* 'pianist' → ***o*** *pianistazinh**o*** 'pianist$_{masc}$+EVAL', ***a*** *pianistazinh**a*** 'pianist$_{fem}$+EVAL'

Differently from other Portuguese suffixes, plural formations in DIM do not correspond with a singular source form, but with a plural inflectional form:

(15) *pão* 'bread' → *pães* 'breads'; *pãozinho* 'bread+EVAL' → ⋆*pãozinhos* (the suitable form is *pãezinhos*)
papel 'paper' → *papéis* 'papers'; *papelzinho* 'paper+EVAL' → ⋆*papelzinhos* (the correc form is *papeizinhos*)

In BP, X-*zinho* is subject to the stress retraction rule applying to phonological phrases. As noted in 16a, the stress clash in 16a causes the first stressed vowel to lose the stress in favor of the immediately preceding syllable (16b). When the final syllable is stressed, in trisyllabic words for example, the secondary stress varies between the first and second syllables (16c):

14 For example, *bambu* 'bamboo' + locative suffix *-al* would create the form ⋆*bamb**ua**l*, with a hiatus between the base and the suffix. This malformed structure is repaired by the epenthetic consonant, generating *bambuzal* 'canebrake'.

(16) a. ca***fé*** ⋆ca***fezi***nho
b. ca***fé*** ***ca***fe***zi***nho
'coffee' 'coffee+EVAL'
c. ani**mal** ***a***nimal**zi**nho ~ a***ni***mal**zi**nho
'animal' 'animal+EVAL'

It can be concluded, therefore, that DIM in BP indeed represent a special case, with features not found in any other morphological process, justifying their special treatment in this chapter. Preference for *-zinho* may be explained by the fact that this form warrants better recognition of the base to which it is associated because it creates its own phonological domain, changing neither the segmental nor the prosodic properties of the Base (↗ 12 Sound-related aspects of Brazilian Portuguese).[15]

4 Derivation and compounding

Compounding is generally understood as a process that combines words or roots to create a morphologically complex unit, whereas derivation requires the presence of an affix, a recurrent bound form. Kastovsky (2009) argues that derivation and compounding, although different, are not always easily distinguishable, because their boundaries are flexible on both sides. Such a claim is based on the existence of non-nuclear entities which can be classified as marginal roots or marginal affixes, in that they exhibit properties of both roots and affixes. In some approaches, such elements were considered to be a different class, between roots and affixes. Such particles have been called affixoids (Marchand 1969). Ten Hacken (2000, 355) notes that the increase in the production of new forms and the reduction of semantic specificity make affixoids resemble affixes; on the other hand, their link to a word makes them resemble roots, as we see below (19), in which *foto* 'photo' compresses the meaning of *fotografia* 'photography' and updates this meaning in the new formations:

(17) fototeca 'collection of photographs'
fotoestúdio 'photo studio'

For our present purposes, the postulation of affixoids highlights the vagueness and malleability of the boundaries between compounding and derivation. In other words,

15 The reader interested in more details on diminutives in BP is referred, besides the works already mentioned, to the following: Skorge (1957), a historical approach to diminutive affixes; Moreno (1997), a study of *-inho*/*-zinho* distribution within the lexical phonology model; Freitas/Barbosa (2013), a variationist investigation of the topic. In the next section, we discuss the differences between compounding and derivation and describe the diversity of WF processes of BP.

affixoids are a valid descriptive entity to segregate the two processes, whether these are of a separate class or not. Such elements indeed stand midway between affixes and words.

Let us now examine two of the main criteria used to distinguish compounding from derivation: (a) the units of a morphologically complex word, and (b) the position where these units occur within the word. It is tacitly understood that compounding involves elements which can appear either on the left or the right of a base. In contrast, affixes are bound forms which conform to rigid positional constraints: for example, prefixes precede the bases, while suffixes categorically succeed them.

The criterion of position confers the status of roots to the units involved in the so-called "neoclassical compounds", since in these morphological constructions there are elements that appear either on the left edge of a complex word or on its right. This is the case with the examples in 18, in which the same morphological unit (in italics) starts the word or precedes the suffix *-ia*, located thus to the right:

(18) *fon*ética/tele*fon*ia '*phon*etics/tele*phony*'
*dem*ocrata/epi*dem*ia '*demo*crat/epi*dem*ic'

Martinet (1979) described as "confixes" those elements without a pre-determined position in word structure. The criterion 'positional mobility', therefore, would lead us to categorize such constituents as roots, since true affixes never change their places. The position, however, is not a criterion which can be considered wholly reliable. Authors like Iordan/Manoliu (1980, 446) use historical arguments to show that the existence of forms with a suffix appearing in the root position is evidence that there was a change in the morphological status of such elements. This was the path taken, for example, by *-metro*, now a productive suffix, although acting as root in words like *métrico* 'metric' and *metragem* 'meterage'.

The noun indicator unit of measurement, *metro* 'meter', came to designate, in complex words, a measuring instrument, accounting for more than 300 new formations today. The vowel /o/ became an integral part of the rightmost element and nowadays it is impossible to assert that *-ômetro* creates technical names of scientific areas, because their formations are quite informal and typical of spoken language. The examples in 19 do not even remotely resemble the old and opaque technicalities (neoclassical compounds) of the 19th century (20), also because the base becomes a word rather than a root:

(19) *beijômetro* 'bad breath detector'
roncômetro 'device that measures a person's snoring during sleep'
bafômetro 'machine that checks if someone has drunk alcohol'

(20) *dinamômetro* 'dynamometer'
ebuliômetro 'ebulliometer'

Bauer (2005) considers the morphological change to be the main empirical basis of his claims here. For him, the boundaries between compounding and derivation are blurred, as many prefixes and suffixes came from roots or words that tended to appear, respectively, on the left or on the right in a morphological construction. Like Kastovsky (2009), we believe that these morphological operations are the prototypical poles of a scale, and thus that some cases are clearly analyzable as compounds or derivatives, and others that are more diffuse, as neoclassical compounds and morphological processes that make use of non morphemic units (the so-called splinters):

To implement a proposal that relates the differences between compounding and derivation, it is necessary to deal with a predetermined set of attributes that applies to emblematic cases (the prototypical ones). In Table 6, we group the main features of the prototypical compounds and prototypical derivatives in BP. Obviously, such differences must be considered as general trends of both processes and not as a form of absolute truth regarding the morphological status of formatives:

Table 6: Main differences between compounding and derivation.

		Compounding	Derivation
Units	I	Words or Roots	Affixes
	II	Free forms or Bound forms which correspond to content words	Bound forms that do not correspond to content words
Structural properties	III	Units with not necessarily fixed position in word structure	Units defined by a predetermined position in a complex word (left or right)
	IV	Roots combine with others morphological types	Suffixes combine almost exclusively with roots
	V	Lexical head is on the left, predominantly	Lexical head is on the right
	VI	Possibility of coordination between the constituents	There is no possibility of coordination between the constituents
	VII	To convey more specific ideas, there is a larger distribution of linguistic units, thus characterizing an open inventory	To convey general ideas, there is a fixed allocation and linguistic units which are not very large, thus characterizing a closed inventory
	VIII	Possibility of inflection between constituents	The inflection is always peripheral
Phonological properties	IX	Expression in more than one prosodic word	Expression in a unique prosodic word
	X	Isomorphism between morphological and phonological word	Absence of isomorphism between these two categories

Table 6: (continued)

		Compounding	Derivation
	XI	Maintenance of segmental and prosodic properties of the bases	Setting the base for applying rules whose domain is the prosodic word
Semantic properties	XII	The units express a lexical meaning	The units update more general semantic content, capable of combining with a larger number of linguistic forms
	XIII	Interpretation often holistic	Interpretation often compositional
	XIV	Can be endocentric or exocentric	Massively endocentric
	XV	Less stable because the meanings of the elements usually undergo metaphorical or metonymic extensions	More stable, presenting predetermined syntactic and semantic functions, defining the possible uses and meanings of derived words
Productivity	XVI	Builds closed sets of non-recurrent words (*ad hoc*)	Builds more complete sets of words (more regular)
	XVII	Presents many manufactured forms	Creates series of words

Except for the criterion (III), all the others may be called into question, leading us to conclude that these criteria really apply to more prototypical elements of the two WF processes here. For example, the difference in (IX) presupposes an isomorphism between the morphological word (MWd) and the prosodic word (PrWd). The prognosis is compatible with most derivative processes, as shown in the following representations, in which brackets indicate PrWds and keys, MWds:

(21) MWd≅PrWd
{[*des* $_{\text{Af}}$ leal $_{\text{Rad}}$] $_{\text{PrWd}}$ } $_{\text{MWd}}$ 'not loyal'
{[*leal* $_{\text{Rad}}$ *dade* $_{\text{Af}}$] $_{\text{PrWd}}$ } $_{\text{MWd}}$ 'loyalty'

Criterion (IX), however, fails in the analysis of prefixes such as *pré-* 'pre-' and *pós-* 'post-', which, undoubtedly, appear in an independent PrWd (22a). Suffixes cited by Booij (2005) as no-cohering, such as *-mente* and *-zinho*, also project PrWds themselves (22b), which avoid the rule of neutralization which applies to pre-stressed mid-vowels (cf. section 2):

MWd≠PrWd

(22) a. {[*prɛ* $_{\text{Af}}$] $_{\text{PrWd}}$ [*pag* $_{\text{Rad}}$ *o*] $_{\text{PrWd}}$ } $_{\text{MWd}}$ 'prepaid'
{[*pɔs* $_{\text{Af}}$] $_{\text{PrWd}}$ [*pag* $_{\text{Rad}}$ *o*] $_{\text{PrWd}}$] $_{\text{MWd}}$ 'paid after'

b. {[*bɛl* $_{\text{Rad}}$ a] $_{\text{PrWd}}$ [*mente* $_{\text{Af}}$] $_{\text{PrWd}}$ } $_{\text{MWd}}$ 'beautifully' (⋆*b*[*e*]*lamente*)
{[*dɔlar* $_{\text{Rad}}$] $_{\text{PrWd}}$ [*zinho* $_{\text{Af}}$] $_{\text{PrWd}}$ } $_{\text{MWd}}$ 'dolar+EVAL' (⋆*d*[*o*]*larzinho*)

The statements in (XII) and (XVII) are not entirely true, according to which affixes necessarily update more general meanings and produce words in series. As shown by Bybee (1985), the relevance of morphological units determines how much the WF scheme in which these elements are instantiated will or will not be applicable on a large scale: the more general the formative meaning is, the more applicable the scheme of which it is a part will be.

Criterion (VI) is also questionable, since it does not apply uniformly to all items classified as affixes in BP. It is impossible, in fact, to delete constituents in coordination in the follow examples (Gonçalves/Andrade 2012):

(23)	des-leal and/or des-honesto	≠	des-leal and/or honesto
	'unfair and/or dishonest'		'unfair and/or honest'
	sabor-oso and/or vali-oso	≠	sabor and/or valioso
	'tasty and/or valuable'		'flavor and/or valuable'

However, there are other complex words, prefixed or suffixed, that allow deletion without restriction of directionality, as in 24. In these data, the deleted elements perform syntactic and semantic functions that are identical to the one that remains, which is the condition necessary to coordination reduction:

(24)	pré-admissão and/or pós-admissão →	pré and/or pós-admissão
	'preadmission and/or post-admission'	'pre- and/or post-admission'
	cordialmente and/or amavelmente →	cordial and/or amavelmente
	'cordially and/or kindly'	'cordial and/or kindly'

None of the properties listed in Table 6 are exclusive to the units under review, or, at least, characterize all the members of these two classes. Neoclassical elements are bound forms and do not behave like words by adding inflection or thematic elements, as other bound roots (*cant+ar* 'sing'; *vid+a* 'life' do, which explains why the following sentence sounds so strange[16]:

(25) *Antropos* são em geral mais peludos que *ginos*.
'Anthropos are generally hairier than gynes'.

16 We are using the invented examples in Portuguese and therefore we do not consider the corresponding Old Greek plurals.

In fact, formatives like *bio-*, *biblio-* and *-teca*, among many others, do not perfectly fit within a root class. As affixes, they are characterized by severe positional constraints, appearing on a specific edge of the word. This is the case with *tele-*, systematically found on the left edge, and *-cida*, which categorically appears on the right:

(26) *tele-novela* 'TV soap opera' *inseti-cida* 'insecticide'
tele-pizza 'pizza by phone' *rati-cida* 'raticide'

Are elements such as *tele-* and *-cida* roots or affixes? The answer to this intriguing question certainly depends on the criterion in focus. Considering the most basic affix properties, i.e., parameters (II) and (III), *tele-* and *-cida* should be recognized, respectively, as a prefix and a suffix, because in addition to being bound forms, they occupy a fixed position in word structure. Conversely, *tele-* and *-cida* do not exhibit the same behavior with respect to the formation of prosodic domains, since only the first projects an independent PrWd (27). In both cases, the front mid-vowel pronunciation should be considered, because the lower-mid vowel ([ɛ]), in *verme* 'worm', becomes upper-mid vowel ([e]), in *vermicida* 'vermicide' (27b), but remains as such in (27a):

(27) a. {[(*t*[ɛ]*le*)]$_{\text{PrWd}}$[(*atendimento*)]$_{\text{PrWd}}$}$_{\text{MWd}}$ not
{[(*t*[e]*le*)[(*atendimento*)]$_{\text{PrWd}}$}$_{\text{MWd}}$ 'remote call services'

b. {[(*v*[e]*rmi*)(*cida*)]$_{\text{PrWd}}$}$_{\text{MWd}}$ not
{[(v[ɛ]*rmi*)]$_{\text{PrWd}}$(*ci.da*)]$_{\text{PrWd}}$}$_{\text{MWd}}$ 'product that kills worms'

Also, in observing criterion (VII) it can be said that the inventory of neoclassical elements is not open as with other roots, and that this once again makes them similar to affixes. It should be noted, finally, that many neoclassical elements have a semantic and syntactic default function, as anticipated by criterion (XV). Thus, we agree with Bauer (2005, 105) when he says that "the label 'neoclassical compound' is then shown to be exocentric, since it is not the case that a neoclassical compound is a compound (under normal readings of the word), but that is a terminological problem rather than a problem of substance".

Therefore, if the uses and meanings of derived words correspond to the functions of affixes, we would have no hesitation in categorizing *-teca* as a suffix, since this formative creates series of words, always providing the same meaning: 'collection'. As shown in 28, all X-*teca* constructions are interpreted compositionally, in accordance with the provisions in (XIII):

(28) *foto-teca* 'photos collection'
esmalto-teca 'nail polishes collection'
xeroco-teca 'collection of photocopies'

If we interpret the differences shown in Table 6 as typical of unquestionably derived or compound constructions, we would have in *peixe-boi* 'manatee' a prototypical example of compounding, and in *saleiro* 'shaker' a clear case of derivation. On the other hand, if we see such differences as attributes to help us in the categorization of WF processes, we would certainly be looking at border situations in examples such as *eletro-choque* 'electric shock', *tiotrocínio* 'sponsored by the uncle', and in this vein why not include *felizmente* 'fortunately' and *pãezinhos* 'biscuits'? Such constructions have attributes that situate them closer to and further from the most exemplary members of these two WF processes.

Considering the existence of a scale between morphological operations, we can better understand the behavior of WF processes difficult to categorize in BP, such as: (a) the truncated combination (*portunhol* 'mixture of Portuguese with Spanish', (b) the sub-lexical replacement (*trêbado* 'very drunk', a word analogically created from *bêbado* 'drunk', with the changing of the /b/, which thus evokes the prefix meaning twice (*bi*), for /tr/, which refers to the prefix meaning three times, *tri*-), and (c) the process known in Brazil as re-compounding (*auto-escola* 'driving school'. These processes are best addressed by the approach advocated here: in BP there are several WF processes that, by sharing compounding and derivation properties. Let us look now at the reasons for this, analyzing first the phenomena referred to as a lexical blend.

A nonconcatenative morphological process often associated with compounding is blending. Although there are two words as inputs for a third (as in compounding), blends differ from compounds because they involve the intersection of bases, instead of concatenation, as in *crentino* (*crente* 'evangelical' + *cretino* 'nitwitted' = 'nitwitted evangelical'.

From a phonological perspective, blends are single PrWds. The output preserves the largest possible number of identical segments of the input, as in *apertamento* 'small apartment' (*aperto* 'tight' + *apartamento* 'apartment' = 'tiny apartment'. Thus, the transition of the first source form to the second coincides with an identical segment or syllable, as in *sa**co**lé* (< *sa**co*** 'bag' + *pi**co**lé* 'popsicle' = a type of popsicle', wherein the bold indicates share segments.

An interesting fact in blend formations is the ability of a non-morphemic piece to engage in new constructions and to acquire morphological status through use. In this case, a phonic sequence is reinterpreted as a morpheme and can become recurrent and create series of words. The formatives in Table 7 are usually combined with parts of words or whole words in BP, and are commonly found in spoken language:

Table 7: Main native splinters in use in contemporary BP.

formative	example	trigger words	meaning of new formations
-drasta	*sogradrasta* 'husband's stepmother"	*madrasta* 'stepmother'	'family on loan'

Table 7: (continued)

formative	example	trigger words	meaning of new formations
-nejo	*pagonejo* 'country music mixed with *pagode*'	*sertanejo* 'born in the backwoods'	'country music with'
-nese	*ovonese* 'mayonnaise salad with hard-boiled eggs'	*maionese* 'mayonnaise'	'mayonnaise salad with'
-tone	*sorvetone* 'ice cream panettone'	*panetone* 'panettone'	'panettone with/of'
-trocínio	*tiotrocínio* 'sponsored by uncle'	*patrocínio* 'sponsorship'	'sponsored by'
caipi-	*caipifruta* 'fruit caipirinha'	*caipirinha* 'a Brazilian cocktail'	'caipirinha made with'
choco-	*chocotone* 'chocolate panettone'	*chocolate*	'chocolate'
piri-	*piricrente* 'overly-extravagant evangelical woman'	*piriguete* 'tart'	'someone is a tart'

The top five particles in Table 7 come from lexical blends. For example, *-nese* is not related to any morphological constituent in *maionese* 'mayonnaise'; it was isolated from the blend *macarronese* 'pasta' + 'mayonnaise' = 'mayonnaise salad with pasta', favoring the creation of words in series by replacing, on the left, the ingredient contained in abundance in the salad made with mayonnaise, that is, *camaronese* 'shrimp' + 'mayonnaise = 'mayonnaise salad with shrimp'. The last four elements are recurring forms derived from clippings (morphological subtractions) which in these cases do not focus on morphemes. In fact, *piri-*, for instance, has no morphological status in the word that originated it, *piriguete* 'tart', whose morphological structure is *pirigu+ete* 'dangerous+fem', form already lexicalized with the elevation of the middle vowel in the base perigo 'danger'.

In the current literature, these particles are called splinters, and are elements occurring in a fixed position, in the same way that affixes do, but they do so because their meanings correspond to roots. Therefore, splinters form a separate class, situated somewhere between roots and affixes. Splinters resemble root or words, but also bear properties of affixes, such as their high lexical production, the fact that they are bound forms, in that they attach at the left edge (*caipi-*) or at the right edge (*-nejo*) in the morphological constructions where they are found. Consequently, they cannot be seen as the result of prototypical compounding. However, they consist of more than one PrWd and are linked to words, evoking the source forms from which they

emerged. This dispenses with the need to analyze them as the result of a derivation process. Therefore, a clear case is seen here for a derivation-compounding cline.

The emergence of a new productive WF schema may happen when speakers start to use a borrowed form, also a splinter, to create series of words. This is indeed what is happening in BP with the use of formatives such as *cyber-*, *-tube-*, and *e-*, which, combined with native bases, form hybridisms like *cyber-avó* 'grandmother well-versed in digital technologies'. Gonçalves/Almeida (2012) note that the use of newly created morphological elements in English from processes such as clipping and abbreviation is becoming common, especially on the Web. For these authors, such a situation—which may seem banal at first glance, as loans in English are quite common in BP—has favored the proliferation of non-native elements in morphological structures. Indeed, since these formatives also adjoin vernacular bases, the neologisms create WF schemes that conform to BP morphological patterns. In the following table, some non-native splinters in use in contemporary BP are illustrated:

Table 8: Main non-native splinters in **use** in contemporary BP.

English splinter	Source form	Meanings	Example
cyber-	*Cybernetics*	'digital'	*ciber-conselho* 'advice on computer'
pit-	*Pitbull*	'aggressive'	*pit-babá* 'aggressive nanny'
-leaks	*Wikileaks*	'leak of information'	*bolso-leaks* 'leak of information about Bolsonaro'
-pédia	*Encyclopedia*	'encyclopedia'	*desciclopédia* 'digital encyclopedia of trivia'
-tube	*Youtube*	'on the computer screen'	*samba-tube* 'learning samba over the Internet'

Another process which, in our view, operates on the boundaries between compounding and derivation is called, in the Brazilian grammatical tradition, re-compounding, a morphological process which creates a compound from a shortening of another compound. In the new formations, however, the base, in a formal metonymy, refers to the meaning of the compound from which it has strayed, moving away from its etymological meaning.

As in English, there are also several formatives in BP that do not fit well within either the class of roots or affixes: these are the aforementioned "affixoids", elements which structure re-compounds. We refer here to particles such as *bio-*, *petro-*, *eco-*, *homo-* and *tele-*, among innumerable others. Perceived as being isolated, they are typical of technical languages (a neoclassical element), and should be treated as formally learned, since they are not products of natural evolution; rather, they have been recovered from classical languages, especially over the last two centuries. However, new formations differ from older technical terms and can come into common use:

(29) *tele*-sexo ‘sex by phone’
eco-turismo ‘ecological tourism’
aero-Lula ‘Lula’s plane’
homo-agressor ‘aggressor of homosexual people’

Undoubtedly, the formations in 29 have morphological elements that actually look like affixes: they (a) are recurring, (b) have a high applicability, (c) are fixed at a specific edge in the word structure, and (d) can be described using identical WF schemes designed by affixes.

If we assume, along with Sandmann (1989), for example, that production in series characterizes derivation, but not necessarily compounding, re-compounding is far from what is expected in prototypical compounding. However, the constituents of a re-compounded word clearly appear in different PrWds, among other factors, by the opening of the mid-vowel in the first formative. In addition, the parity between the shortened form and the full form suggests a compounding process. Finally, a kind of syntactic structure can enhance non-fulfilment of a first base, when two complex words are placed in parallel, such as in 30:

(30) *tele* e *auto*-atendimento ‘phone service and self-service’
foto e *tele*-jornalismo ‘photojournalism and television journalism’

The behavior of re-compounds is similar to formations in *-mente* (*livre e continuamente* ‘freely and continuously’, which would lead us to claim, for this and other reasons, that X-*mente* constructions are not a prototypical case of derivation. It can be concluded that the boundary between compounding and derivation is diffuse, at least in BP. The idea of a scale is promising, since it accounts for the typological heterogeneity of the Portuguese WF system, which allows for words, roots (native and neoclassical), splinters (natives and non-natives), different kinds of affixes, and affixoids as morphological units in a wide variety of WF processes. A more systematic study of the diversity of WF processes in BP can be found, for example, Andrade (2013) and Gonçalves (2016). Rio-Torto (2013) brings a monumental study on WF in EuP.

5 Other nonconcatenative word-formation processes

BP has a wide variety of nonconcatenative WF processes. Apart from blends, as discussed above, there are shortening processes (clipping or truncation), reduplication patterns and formation of acronyms, including reverse acronyms. These phenomena differentiate BP from other Lusophone varieties (Villalva/Gonçalves, 2016). This section devotes some attention to these processes, typical of spoken language, citing the most common examples in each case.

Clipping is quite common in BP. It is a mechanism by which a word is shortened without its lexical meaning being affected, but with frequent stylistic or pragmatic nuances. It eliminates phonological material at the right periphery of the base. Clippings may (31a) or may not (31b, 31c) affect morphological constituents:

(31) a. prolet+ário → *prolet*+a 'proletarian'
b. cervej+a → *cerv*+a 'beer'
c. bijuteri+a → *biju* 'jewelry'

The patterns exemplified above require morphological and prosodic information. In 31a, we find a word formed by a root base and the thematic index *-a*, a constituent unrelated to the gender of the base (the products are not previously specified to gender). In 32b, the base root is not fully present in the truncated form, but, as in 31a, the clippings are stressed on the penultimate syllable, always forming a trochee at the right edge of the shortened form. In these two groups, the affixation of the thematic index (*-a*) always occurs, but not in 31c. In this case, the two first syllables of the base are kept, which form an iambic foot. The first pattern of clipping—the most common one—can also affect compounds, as in *grã-fino* → *granfa* 'stylish' and *São Paulo* → *Sampa* 'Brazilian city'.

Forms such as *agro-*, *eletro-* and *foto-*, which are neoclassical roots, often become words through clipping. As words, they can vary in number, as in *A casa tem duas* ***hidros*** 'The house has two tubs' and *Tenho dois amigos* ***homos*** 'I have two homosexual friends'.

Reduplication is another BP nonconcatenative process. It is a mechanism of very limited productivity in EuP, but in BP new reduplicated forms are easy to find. For example, the addition of a word-final VCV template expresses intensification:

(32) chor+o chororô 'crying/excessive crying'
bol+o bololô 'group of people/confusion'
baf+o bafa'fá 'breath/quarrelling'

In all the words that express intensity through the use of this strategy, the resulting vowels are always identical, and thus there is perfect harmony in the vowel melody. Examples such as those in 32 lead us to consider that the copying process makes use of the melodic elements of the root (not the word), since the thematic index of the input never emerges in the output. Thus, from *ch[o]ro* we get *chororô*, with an upper-mid vowel [o] as the nucleus of the three syllables in the morphological output. The same can be said of *b[a]fo*, in which the elimination of the thematic vowel produces three identical low vowels (*bafafá*).

This pattern of reduplication appears in other cases, and always expresses intensity, as in *bafafá* 'excessive verbal confusion' and *sururu* 'clutter', for which it not always easy to designate a base (*bafo* and *suruba*, respectively 'whiff' and 'bacchanal'.

For such reduplicative words, we can define an abstract phonological template, C(C) $V_iC_jV_iC_jV_i$, where the i and j subscripts indicate full feature identity.

The second subtype, again more common in BP than in other varieties (Villalva/Gonçalves, 2016), involves reduplication of the verb to form a ViVi compound. These forms can convey two meanings: an action (33a) or an object (33b). In some cases, both meanings can be observed in the same word (33c):

(33) a. *borra-borra* 'smudge-smudge = repeated smudging'
b. *bate-bate* 'hit-hit = dodgem cars'
c. *pula-pula* 'jump-jump = act of jumping repeatedly/a trampoline'

The bases of ViVi compounds are generally disyllabic. There are also cases like *em-**pur**ra-empu**r**ra* 'push-and-shove', with three syllables, but these always start with an onset-less syllable. Since the reduplication of the verb base is governed by prosodic requirements, the reduplications always end in open syllables. Finally, the main morphological feature of this process is the selection of the third-person present singular indicative: we assume that this is the unmarked form of the verb paradigm, which allows for the reinterpretation of the verb as a noun.

Acronyms are widely created in current BP. A defining characteristic of acronyms, unlike other non-morphemic processes (blending, clipping),[17] is the fact that they are widely found in the written language. A new phenomenon, identified in recent literature (Fandrych 2008), is the increase of so-called reverse acronyms, in which the created forms are based on a commonly used acronym which is thus reinterpreted. Staring with the letters, a search is made for those words that represent the new idea that needs to be expressed. Ironic intentions are the driving force behind the following playful reinterpretations:

(34) MMA – Mixed Martial Arts
Monte de Machos se Agarrando
lots of males hugging
'a group of males hugging'
SUS – Sistema Único de Saúde
system unified health
Sistema Único de Sacanagem
meanness
'unified system of meanness'

17 According to Marchand (1969, 1), these formations are not morpheme-based. For Fandrych (2008), some recent studies agree with Marchand, calling non-morphemic WF processes "unpredictable" or even labelling them as "oddities". This chapter will not go into this discussion.

Of all the nonconcatenative WF processes, the most important, due to the role they have played in contemporary BP morphology, are blending and clipping, due to the creation of new morphemes (splinters), as we have seen. Other discussions of nonconcatenative WF processes in BP can be found in Araújo (2000), Gonçalves (2004), and Basilio (2005).

6 Conclusion

For the description of Brazilian morphology here, not only have we focused on the more specific uses, ones which are often poorly described in the traditional literature, but we have also been careful to characterize different borders in morphology, analyzing a wide variety of morphological processes, such as clipping, blending, compounding and acronyms. We have looked specifically at nonconcatenative processes, which are not recognized in European Portuguese, as well as pragmatic aspects (such as the pejorative DIM).

In conclusion, we suggest that the facts as described point to BP as now being a language with a greater tendency to synthesis than other Lusophone varieties.

7 References

Altakhaineh, Abdel Rahman (2014), *The Interaction Between Inflection and Derivation in English and MSA: An Insightful Glimpse Into the Boundaries*, Chisinau, Lambert Academic Publishing.

Andrade, Katia Emmerick (2013), *Proposta de continuum composição-derivação para o português do Brasil*, Doctoral thesis, Rio de Janeiro, UFRJ.

Araújo, Gabriel (2000), *Morfologia não-concatenativa: os portmanteaus*, Campinas, Cadernos de Estudos Lingüísticos 39/1, 35–45.

Aronoff, Mark/Fudeman, Kristen (2005), *What is Morphology?*, Hoboken, New Jersey, Wiley-Blackwell.

Basilio, Margarida (1987), *Teoria Lexical*, São Paulo, Ática.

Basilio, Margarida (2005), *Fusão Vocabular como Processo de Formação de Palavras*, in: *Anais do IV Congresso Internacional da ABRALIN*, Rio de Janeiro, UFRJ, 39–45.

Bauer, Laurie (2005), *The borderline between derivation and compounding*, in: Wolfgang Dressler et al. (edd.), *Morphology and its Demarcations*, Amsterdam/Philadelphia, Benjamins, 97–108.

Bisol, Leda (2010), *O diminutivo e suas demandas*, D.E.L.T.A. 26/1, 59–85.

Booij, Geert (2005), *Compounding and derivation: evidence for Construction Morphology,* in: Wolfgang Dressler et al. (edd.), *Morphology and its Demarcations*, Amsterdam/Philadelphia, Benjamins, 109–132.

Brown, Penelope/Levinson, Stephen (1978), *Universals in language usage: Politeness phenomena,* In: Edward Goody (edd.). *Questions and politeness*, Cambridge, Cambridge University Press, 56–324.

Bybee, Joan (1985), *Morphology: a study of the relation between meaning and form*, Cambridge, Cambridge University Press.

Bybee, Joan (2010), *Language, usage and cognition*, Cambridge, Cambridge University Press.
Câmara Jr., Joaquim Mattoso (1970), *Estrutura da língua portuguesa*, Petrópolis, Vozes.
Colnaghi, Carmen (2006), *Flexão e derivação: um labirinto gramatical*, Passo Fundo, UPF.
Fandrych, Ingrid (2008), *Submorphemic Elements in the Formation of Acronyms, Blends and Clippings*, Lexis – E-Journal in English Lexicology 2, 208–236.
Freitas, Myrian Azevedo/Barbosa, Maria Fernanda (2013), *A alternância do diminutivo "-inho/-zinho" no português brasileiro: um enfoque variacionista*, Alfa 57/2, 577–605.
Gonçalves, Carlos Alexandre (2004), *Processos morfológicos não-concatenativos: formato prosódico e latitude funcional*, Alfa 48/2, 30–66.
Gonçalves, Carlos Alexandre (2005), *Flexão e derivação em português*, Rio de Janeiro, UFRJ.
Gonçalves, Carlos Alexandre (2011), *Iniciação aos estudos morfológicos*, São Paulo, Contexto.
Gonçalves, Carlos Alexandre (2016), *Atuais tendências em formação de palavras*, São Paulo, Contexto.
Gonçalves, Carlos Alexandre/Almeida, Maria Lucia Leitão de (2012), *Por uma cibermofologia: abordagem morfossemântica dos xenoconstituintes em português*, in: Maria Mollica et al. (edd.), *Linguística e Ciência da Informação*, Curitiba, Appris, 105–127.
Gonçalves, Carlos Alexandre/Andrade, Katia (2012), *El status de los componentes morfológicos y el continuum composición-derivación en portugués*, Lingüística 28/2, 119–145.
Iordan, Iorgu/Manoliu, Maria (1980), *Manual de lingüística románica*, Madrid, Gredos.
Kastovsky, Dieter (2009), *Astronaut, astrology, astrophysics: about combining forms, classical compounds and affixoids*, in: Roderick W. McConchie et al. (edd.), *Selected Proceedings of HEL-LEX 2*, Somerville, Cascadilla Proceedings Project, 1–13.
Langacker, Ronald W. (1987), *Cognitive linguistics, Foundations of cognitive grammar*, vol. 1: *Theoretical prerequisites*, Stanford, Stanford University Press.
Lee, Seung-Hwa (1997), *Sobre a formação de diminutivos no português do Brasil*, Belo Horizonte, UFMG.
Levinson, Stephen (1983), *Pragmatics*, Cambridge, Cambridge University Press.
Marchand, Hans (1969), *The Categories and Types of Present-day English Word-formation*, München, Beck.
Martinet, André (1979), *Grammaire fonctionelle du français*, Paris, Didier.
Moreno, Cláudio (1997), *Morfologia nominal do Português: um estudo de fonologia lexical*, Doctoral Thesis, Porto Alegre, PUCRS.
Pereira, Eduardo Carlos (1908), *Gramática expositiva*, Rio de Janeiro, TPS.
Rio-Torto, Graça Maria (2013). *Gramática derivacional do português*, Coimbra, Coimbra University Press.
Rocha, Luis Carlos de Assis (1994), *Flexão e derivação*, Belo Horizonte, UFMG.
Rosa, Maria Carlota (2000), *Introdução à Morfologia*, São Paulo, Contexto.
Rosch, Eleanor (1978), *Principles of categorization*, in: Eleanor Rosch/Barbara Lloyd (edd.), *Cognition and categorization*, Hillsdale, Erlbaum, 7–36.
Sandmann, Antônio José (1989), *Morfologia Lexical*, São Paulo, Contexto.
Silva, Augusto Soares (2006), *O mundo dos sentidos*, Lisboa, Fundação Calouste Gulbenkian.
Skorge, Silvia (1957), *Os sufixos diminutivos em português*, Boletim de Filologia, 16–17/3–4, 40–90.
Spencer, Andrew (1998), *Introduction*, in: Andrew Spencer/Arnold Zwicky (edd.), *The Handbook of Morphology*, Oxford, Blackwell, 1–10.
Ten Hacken, Pius (2000), *Derivation and Compounding*, in: Geert Booij et al. (edd.), *Morphologie—Morphology*, Berlin/New York, De Gruyter, 349–360.
Villalva, Alina (2000), *Estruturas morfológicas: unidades e hierarquias nas palavras do português*, Lisboa, Fundação Calouste Gulbenkian.

Villalva, Alina/Gonçalves, Carlos Alexandre (2016), *The phonology and morphology of word formation*, In: Leo Wetzels/João Costa/Sergio Menuzzi (edd.), *The Handbook of Portuguese Linguistics*, Oxford, Wiley-Blackwell, 167–187.

Vivas, Vitor de Moura (2015), *Abordagem de padrões derivacionais nas marcas de modo-tempo-aspecto e número-pessoa*, Doctoral thesis, Rio de Janeiro, UFRJ/Faculdade de Letras.

Sanderléia Roberta Longhin and Maria Clara Paixão de Sousa

14 Syntax

Abstract: This chapter provides an outline of contemporary Brazilian Portuguese syntax. For an overview of the diachronic changes that led to the emergence of this syntactic system, we refer the reader to ↗6 Historical syntax. The outline offered in the present chapter is informed by the perspectives of two different theoretical and methodological frameworks: the functionalist perspective, which privileges the relations between language and its social functions, on levels that go beyond the sentence, and the formalist perspective, in which syntax is seen as the grammatical-internal level at which core linguistic properties are computed as a means of relating sound and meaning.

Keywords: syntax, pronominal system, predication, complex predicates, relativization strategies

1 Presentation: functionalist and formalist views of syntax

"Syntax", in its most basic definition, is the study of the rules and patterns by which a language combines words to build larger meaningful units such as phrases, clauses and sentences, that is, rules and patterns that can be used in a dynamic way, producing infinite results from a finite set of resources (see among others Baker 2002). Not surprisingly, the study of the "syntax" of any given language will privilege different aspects of these patterns of sentence-formation, depending on how one understands "language" in the first place, and, given that the purpose of this chapter is to discuss specific syntactic aspects of Brazilian Portuguese (henceforth BP) from both a functional and a formal viewpoint, some introductory remarks on how "syntax" is viewed by each of these approaches are necessary.

On the functionalist view, the structural settings of languages are considered to be determined by the functions which they fulfill in the life of social individuals. From this perspective, the relations between form and function are naturally inseparable and yet unstable. *Inseparability* implies basic functionalist assumptions, such as the non-autonomy of the linguistic system in regard to cognitive and sociocultural forces, the weight of contexts of use in linguistic descriptions, and the relationship of mutual determination between the syntactic, semantic and pragmatic components. *Instability* implies the system's continuous and progressive reorganization, as evidenced by the non-discrete nature of categories and processes and the coexistence of stable usage patterns and emerging patterns.

The formalist view of syntax is largely represented today by the generative paradigm. Generative linguistics is interested in the study of an abstract object which it calls

https://doi.org/10.1515/9783110405958-015

"Grammar", rather than in the study of languages as social, historical facts. Grammar is, accordingly, an inaccessible, logical object that produces the specific linguistic facts we observe, and generative linguistics tries to understand Grammar through the theoretically informed observation of languages. For instance, from this perspective the great variety of languages that have emerged historically is explained by the proposal that Grammar involves not only immutable principles, but also variable parameters; it follows from this that, by studying how different historical languages function—how widely they vary, and, conversely, how closely they resemble each other—we will be able to understand the principles and parameters of Grammar (and, ultimately, how the unique human property of the "faculty of language" works). Syntax is a central component in the generative framework, as it involves the core grammatical mechanisms that relate sound and meaning in different ways in different historical languages.

The formalist and the functionalist views of syntax, then, are very distant in terms of their theoretical frameworks and their objects of study. In this chapter, however, we attempt to summarize those aspects of BP that have been of particular interest in both these approaches to syntax. We are assisted here by the fact that much of our exposition will be grounded in the descriptions of spoken BP developed by the project *Gramática do Português Falado*, 'Grammar of Spoken Portuguese', conducted by linguists from several Brazilian universities in the 1980s and 1990s, and coordinated by Ataliba T. de Castilho. The project pioneered the combining of work from both formalist and functionalist syntacticians, many of whom will be referred to in the present study. The results of this very intensive research initiative have been published in their entirety in successive editions of the *Coleção Gramática do Português Falado*, 'Collection Grammar of Spoken Portuguese', with Volume II (Kato/Nascimento 2015) presenting syntactic research conducted by the formalist team, and Volumes III and IV (Ilari 2014; 2015) presenting the syntactic research by the functionalist team.

The present chapter address those topics that are relevant to both perspectives in a broad sense, from more morphosyntactic issues to questions of sentential syntax, organized as follows: the pronominal system (Section 2), predication strategies (Section 3), and relativization strategies (Section 4). We will begin each section with a review of functionalist work, and follow it with a review of formalist work, seeking thus to reveal the synergy between the two approaches wherever this is possible. Hence we will offer an outline of the syntax of BP that can serve as a basis for further research from both theoretical perspectives.

2 Pronominal system and verb inflection

Among the linguistic features that characterize BP and make it stand out from other varieties of Portuguese, its pronominal system, and by association its verbal inflection paradigm, have probably been discussed more widely and for longer than any other. Indeed, the very first descriptions of the variety of Portuguese spoken in Brazil referred to

this (Coelho 1880, among others), which would become a central point of debate for those first 20th-century linguists concerned with the description of "the language of Brazil" (Neto [3]1976; Melo [3]1975), and for whom the particularities of the verbal-inflection paradigm in the "popular" varieties of what we now call BP were a source of much debate (see Paixão de Sousa 2010, for a historical account of these discussions). In more recent research, the focus on verbal inflection and the pronominal system remains. To begin with, it has been observed widely that "The set of personal pronouns that is effectively used in Brazilian Portuguese is very different from that for European Portuguese, not only in the subject position (nominative), but also for direct (accusative) and indirect (dative) complements, and for the possessive (genitive)", as noted in ↗6 Historical syntax, where a detailed account of important aspects of diachronic developments leading to such differences is outlined. Here, we focus on the main consequences of this change in BP, in that the rearrangements in the personal pronoun system of BP allow us to understand aspects of its current composition and form of functioning.

In this sense, the two key points here will be pronouns and verb forms. Consider, in Table 1 below, a simplified picture of Present-day BP's verb-pronoun paradigm, in the subject position.

As we can see, in Old Portuguese and in Present-day European Portuguese, the paradigm includes six pronominal forms and six corresponding verbal forms, whereas in BP, the picture is more varied. The most immediate observations are:

i. **Pronouns.** In BP there are two "new" forms, which we analyze as pronominal forms—respectively, the second person singular-plural pair *você-vocês*, and the first person plural form *a gente*.

ii. **Verb forms**. There is considerable variation in the verb forms used with the second person (singular and plural) pronominal forms and with the first person plural pronominal forms:
- With the second person singular pronominal *tu*, the verb may be second person singular or third person singular (*tu falaste ~ tu falou*). For the new pronominal form *você*, the verb is consistently third person singular (*você falou*); with the new second person plural pronominal *vocês*, the verb is consistently third person, but varying between plural or singular (*vocês falaram ~ vocês falou*).
- With the first person plural pronominal *nós*, the verb may be first person plural or third person singular (*nós falamos ~ nós falou*); with the new first person plural pronominal *a gente*, the verb form may be first person plural or third person singular (*a gente falamos ~ a gente falou*)[1].

1 There is also the following variation in the form of the second person plural verbal inflection: *falamos ~ falamo*, 1PL speak.PRES.1PL; the two forms can appear with *nós* and with *a gente*. We consider this a morpho-phonetic variation, so that both forms correspond to the same grammatical person and number; hence this variation is not shown on the table describing the paradigm.

Table 1: Paradigm—Old Portuguese, European Portuguese and Brazilian Portuguese.

Old Portuguese, European Portuguese	*Brazilian Portuguese*				
Eu fal*ei* 1SG speak. PST.1SG	**Eu** fal*ei* 1SGspeak. PST.1SG				*I spoke*
Tu falas*te* 2SG speak. PST.2SG	**Tu** falas*te* 2SG speak. PST.2SG	~ **Tu** fal*ou* 2SG speak. PST.3SG	~ **Você** fal*ou* 2SG speak. PST.3SG		*you spoke*
Ele fal*ou* 3SG speak. PST.3SG	**Ele** fal*ou* 3SG speak. PST.3SG				*he spoke*
Nós falam*os* 1PL speak. PST.1PL	**Nós** falam*os* 1PL speak. PST.1PL	~ **Nós** fal*ou* 1PL speak. PST.3SG	~ **A gente** falam*os* 1PL speak. PST.1PL	~ **A gente** fal*ou* 1PL speak. PST.3SG	*we spoke*
Vós falas*tes* 2PL speak. PST.2PL	**Vós** falas*tes* 2PL speak. PST.2PL	~ **Vocês** falar*am* 2PL speak. PST.3PL	~ **Vocês** fal*ou* 2PL speak. PST.3SG		*you spoke*
Eles falar*am* 3PL speak. PST.3PL	**Eles** falar*am* 3PL speak. PST.3PL	~ **Eles** fal*ou* 3PL speak. PST.3SG			*they spoke*

A great deal of research has been conducted to explain the BP paradigm shown above regarding these two main points (the rise of two new pronominal forms, and the associated inflection patterns), both diachronically and synchronically; and, as we will see, the two perspectives are closely interwoven. For a thorough discussion of the diachronic progress of this change, see ↗6 Historical syntax (Section 2.1) and references therein. Here, we will focus on the association between the introduction of the new pronominal forms and the variation in the verb-subject agreement paradigm. The first and most notable contribution in relation to this, from a diachronic perspective, is that it explains the use of third person verb forms with the "new" pronominals. The second person pronominal forms *você*/*vocês* emerge from the reanalysis of Old-Portuguese forms of address *Vossa Mercê* 'your grace'. In fact, these are still common to some other varieties of Portuguese, and the innovation as far as BP is concerned is their status as pronominals. This explains the use of third person verb inflection with these forms—a trivial fact when they are forms of address, which remained as a part of their use as pronominals. As for the new first person plural form *a gente*, its history may also explain the use of third person verb forms, but in a slightly different way.

The form *a gente* also derives from a nominal form, in this case not a form of address but rather the collective and indeterminate noun *gente* 'people'.

The use of third person forms with the new pronominals, however, is not the only striking feature of this paradigm in BP: there is also variation in the verb forms within each new form, and in fact also with old forms. As regards variation within the new forms, note that the second person plural pronominal *vocês* may appear with a third person singular verb (*vocês falou*) as well as with plurals (*vocês falaram*); and the first person plural pronominal *a gente* may appear with a third person singular verb (*a gente falou*) as well as with first person plural (*a gente falamos*). With the old pronominals there is also variation in the use of verb forms: the second person singular *tu*, the first person plural *nós*, and the third person plural *eles* may all appear with third person singular verb forms (*tu falou*, *nós falou*; *eles falou*), as well as with second person singular, first person plural, and third person plural forms, respectively (*tu falaste*, *nós falamos*; *eles falaram*).

Our summarized paradigm in Table 1, therefore, shows an ample range of variation: between singular and plural verb forms, between second and third person verb forms, and between the different pronominal forms themselves (as the old forms *nós* and *tu*/*vós* still co-exist with the new first person plural form *a gente* and the new second person forms *você*/*vocês*, respectively). However, it is crucial to notice that this variation is not as unconditioned as the simplified picture we have painted might suggest: on the contrary, it is marked by well-known sociolinguistic conditioning factors, and some of the forms are very strongly stigmatized socially. This variation is at the heart of the analysis of the socio-historical factors leading to the formation of different socio-dialects in spoken BP, as Lucchesi (2001) has shown. In BP today, real language use in various written and spoken domains suggests that a part of this system is still undergoing change, and that these changes affect different sociolinguistic groups in different ways;[2] moreover, within each group, parameters related to the speech register, phonic salience of verb forms, and the degree of indetermination of the constructions determine the predominance of each form (see Lopes 1999; 2003; Ilari 2014, in particular for the conditioning factors of the use of *a gente* vs. *nós*).

2 Notice that, as the range of possible socio-linguistic variations is very wide, at one extreme of such variation there is the possibility of a paradigm where the only verb-person contrast is between first person singular and all other persons (*eu falei*, *você falou*, *ele falou*; *a gente falou*, *vocês falou*, *eles falou*). This is in fact amply documented in so-called "popular" BP, i.e., the variety of the language most closely associated with sociolinguistic parameters such as "oral", "informal", and, most importantly, "low formal education of speakers". Thus it constitutes a marker of social stigma. However, note that in varieties associated with "higher education of speakers" (thus not marked with social stigma) the paradigm may also be very heavily marked by the third person, the only difference being the plural verb forms used with *vocês* and *eles*: *eu falei*, *você falou*, *ele falou*; *a gente falou*, *vocês falaram* and *eles falaram*.

In view of the breadth and importance of the debate around the conditioning factors leading to the rearrangement in morphosyntax of the personal pronoun system in BP, we have chosen here to highlight one central point: *the generalization of the usage of third person singular form*. In effect, this point has been taken as crucial in both functionalist and formalist discussions. Studies from the functional perspective have long suggested that this structural setting has introduced the third-person verb form neutralization (*ele*/*ela estuda* 'he/she studies', *você estuda* 'you study', *a gente estuda* 'we study') in BP—a phenomenon which, in turn, goes towards filling the subject position as a means of avoiding ambiguity, since the pronouns gain the status of being only person markers. For a careful and detailed account of this process in the light of the historical emergence of the vernacular forms *você*/*a gente*, see Faraco (1996). In addition, neutralization here is related to the fact that the pronouns *você* and *a gente* may have a general, indeterminate character:

(1) engraçado que **você** saindo do Brasil...
funny that **you** leave.GER of-the Brazil...

a gente sente uma falta muito grande dessa parte de verduras
we feel.PRES.3SG a loss very big of-these part of vegetables
(Ilari 2015, 35)

'it is funny that when ones leaves Brazil...
one misses this kind of vegetables very much'

This use of the new forms *você*/*a gente* as indeterminates has also been of interest in generative studies. Within this framework, the restructuring of the pronominal system of BP has been addressed in close relation to the loss of the *null subject parameter*, one of the most notable topics of current research into BP in this field. The loss of null subjects, from this perspective, is connected not only to the particular pronominal paradigm of BP, but also the properties of this grammar regarding the expression of arguments in a more general sense, including the order of constituents, for instance, as we will discuss in Section 3 below. The contrast in the BP paradigm regarding verbal inflectional morphology is also a central issue in generativist research relating to the grammatical restructuring around the expression of subjects, just as it is in functionalist research. Here, though, the "cause and consequence" relation is less clear, in that opinion is divided on one central question: is the restructuring of the pronominal system a result of changes in inflectional morphology, or, on the contrary, are the changes in inflectional morphology caused by the restructuring of the pronominal system? An extensive review of this issue and its consequences for the analysis of the loss of null subjects can be found in Torres-Morais (2001).

This tendency of showing less agreement morphology than in other varieties of Portuguese (either Old or European Portuguese), which we described here briefly in terms of the subject-verb paradigm, is a central morphosyntactic characteristic of BP.

This characteristic is related to other important aspects of its syntax, both within the verbal paradigm (particularly regarding personal infinitives) and elsewhere (particularly agreement morphology within the noun phrase, with plural forms tending to be marked on the determiner but not on the noun in some dialects: *as meninas ~ as menina*, i.e., 'the.PLURAL girl.PLURAL ~ the.PLURAL girl.SINGULAR'). For a description of this aspect of BP morphosyntax, see Castilho (2010, 457–461).

Finally, it is important to note that the new pronominal forms, both for second and first person, are present not only in the nominative case (as shown in Table 1 above), but also in accusative, dative and oblique forms (*ele chamou a gente* 'he called us'; *ele deu o cartão pra gente* 'he gave us the card'); once again, ↗6 Historical syntax provides a detailed account of their rise in positions other than the subject. Here we will focus on their role as subject pronominals, as this constitutes a very intense area of debate on the syntax of BP, and leads to significant observations regarding other aspects of sentence construction, as we will see.

3 Predication: the expression and order of phrasal constituents

The complex and hierarchical relation formed by the verb with the arguments that complete it in order to form sentences is a crucial factor in the description of a language. We will examine here some instances of this relation that help us to outline the syntax of BP in the verbal domain, taking into account some particularly interesting work in this respect, such as Galves (2001); Neves (2006); Castilho (2010); Ilari (2014); Berlinck/Duarte/Oliveira (2015). To begin with a more general account of this aspect of syntax, we may note that in BP, verbs may require one to three argumental constituents, as in (2) to (4), whereas those which do not require any argumental element (verbs referring to climatic phenomena) or which require four elements are considered exceptional, as shown in (5) and (6), respectively (all examples below are from Ilari 2014, 88):

(2) Eu$_1$ acordo V1
I wake-up.PRES.1SG
'I wake up'

(3) Eu$_1$ conheço Bernadete$_2$ V2
I know.PRES.1SG Bernadete
'I know Bernadete'

(4) Ele$_1$ não me$_2$ deu o violão$_3$ V3
he not me give.PST.3SG the guitar
'He did not give me the guitar'

(5) Nevou em Gramado V0
snow.PST._ in Gramado
'It snowed in Gramado'

(6) Eu$_1$ traduzi um trecho$_2$ do inglês$_3$ para o português$_4$ V4
I translate.PST.1SG a passage from-the English into the Portuguese
'I translated a passage from English into Portuguese'

The requirement to have arguments is not, of course, an exclusive property of verbs, and is also seen in many nouns and adjectives as well as some prepositions, as shown in (7) to (9) below (examples from Ilari 2014, 104):

(7) A estrada não estava **pronta** ainda
the road not be.IMPERF.3SG ready yet
'The road was not ready yet'

(8) Itauçu que é uma **cidadezinha** lá que
Itauçu which be.PRES.3SG a town over there which
inclusive me ofereceu hospedagem
even me offer.PST.3SG housing
'Itauçu, which is a town over there which even offered me housing'

(9) Foi uma palestra de oito minutos **sobre** a ocupação
be.PST.3SG an lecture of eight minutes about the occupation
da Amazônia
of-the Amazon
'It was an eight-minute lecture about the occupation of the Amazon'

In (7), *a estrada* 'the road' is the argument of the adjective *pronta* 'ready'; in (8), the argument of *cidadezinha* 'small-town' is *Itauçu*; in (9), *a ocupação da Amazônia* 'the occupation of the Amazon' and *uma palestra de oito minutos* 'an eight-minute lecture' are the arguments of the preposition *sobre* 'about'. In those cases, the verbs (respectively, *estava* 'be.IMPERF.3SG', *é* 'be.PRES.3SG', *foi* 'be.PST.3SG') are not responsible for the argument structure; they are required by syntax since nouns and prepositions cannot construct sentences on their own. The predicative nature of nouns and prepositions, along with the particularities of the adjoining verbs' syntactic and semantic way of functioning (*ser* 'be', *estar* 'be', *ficar* 'stay', *tornar-se* 'become', *andar* 'go' etc.) open up a hitherto little explored field of study (see Neves 2006; Ilari 2014).

From a functionalist point of view, the predicate construction shows the interface between the various levels of linguistic analysis, since it consists of an inherently grammatical fact which finds a pragmatic motivation and is subject to semantic restrictions. In this interface, the various ways of realization of a predicate are not random choices, but rather result from constraints related to communicative aims. Departing from the central position of the *verb* in the predicate construction, the number and nature of the constituents which will complement the verb, as well as the means of expression and ordering of these constituents, are determined primarily by semantic and pragmatic factors.

In generative syntax, the sentence is defined as the projection of the argument structure of a verb, that is, as the projection of the verb's semantic selection properties. Thus, the construction of predicates is also of central interest here. However, it is not seen as being linked to pragmatic motivations; instead, it is the verb's formal traces, pre-defined in each linguistic system, that determine the argument structure that it may project, and, thus, those pre-defined formal traces are at the base of the projection that forms the sentence. In this context, BP has received particular attention due to what some researchers have called its "special predication properties" (Galves 2001), this observed in the characteristics of the expression of subjects, as we will see.

In subsections 3.1 and 3.2, we analyze the verb as a predicate matrix to describe features of the constituents' modes of expression and ordering in BP. In subsection 3.3, we look at constructions in which the verbs are not predicators.

3.1 Phrasal constituents expression

Although both lexical and null subjects are possible in BP, Duarte (1993), Neves (2006), Castilho (2010), among others, observe changes which point to the increasing tendency towards filling the subject. An apparently decisive factor in this increase is the neutralization of the verbal morphology in the different grammatical persons as a result of changes in the personal pronoun paradigm, this through the process that we summarized in Section 2 above. Other factors involved in the expression or non-expression of the subject include the syntactic and semantic type of the verb, polarity, subject co-reference in coordinate and subordinate constructions, personal infinitive, as well as prosodically determined factors (Neves 2006; Castilho 2010).

In functionalist studies, the way constituents are expressed in the argument structure is seen to be motivated by pragmatic-discoursive factors, such as the information flow, the topic organization of utterances, and the need to provide a descriptive specification of the referents (Neves 2006). Therefore, the expression of subjects or complements by means of a noun phrase, a pronoun, zero (ellipsis), or clause represents syntactic patterns guided by pragmatic choices.

As to the expression of the subject, Neves (2006) argues in favor of the weight of information processing, suggesting that its realization through noun phrases is

strongly related to the need to describe an entity as being *new*, whereas leaving the subject blank, which is possible in BP, could involve shared information. Apart from the relevance of the informational status, functionalist and variationist-oriented studies (see Castilho 2010, among others), when considering the semantic properties of the subject and the argument structure of the verb, provide evidence that the agentive character of the subject and the rich verbal morphology are factors which support the omission of the subject.

In formalist studies, the expression of the subject is, perhaps, the most salient and well-studied aspect of BP grammar. The observed tendency towards the lexical expression of subjects is seen as evidence that BP is the result of a grammatical change in relation to other varieties of the language (diachronic varieties, such as Old and Classical Portuguese, and synchronic varieties, such as European Portuguese). This is particularly relevant since the property of a language in terms of allowing or not allowing the subject to be expressed as *null* is considered to be a core property that differentiates grammars; in other words, a *parameter*. It has largely been considered that BP has lost the *null-subject parameter*, and there is a wealth of research into this process and its consequences for BP grammar, see particularly the studies collected in Kato/Negrão (2000) and Galves (2001). In order to understand the relevance of this issue for generativist research, it is vital to note that in formalist syntax, *null subjects* are conceptualized as empty grammatical categories, that is, as constituents that may have semantic and formal content, even though they have no *phonological* content (i.e., even though they are not pronounced). In other words, on this view, even if it seems that a subject is "not there", it *is* there, since it carries both formal properties (for instance, it might be interpreted as "third person singular") and semantic properties (for instance, as an argument of verbs, it carries thematic role interpretation, such as "agent"). According to generative syntax, in order for a grammatical category to be "licensed" as null, a group of conditions must apply that will ensure its interpretation and the felicity of the constructions in each language. This can be seen in the examples below. Here (and throughout the chapter) we represent empty categories with a gap sign, __, and as *ec* in the gloss, placing them where we claim their interpretable position in the structures is:

(10) Null subjects in Brazilian Portuguese

a. Ninguém sabe o que __ quer
nobody know.PRES.3SG the what *ec* want.INF
'Nobody knows what they want'

b. O Pedro disse que __ ia viajar
the Pedro say.PST.3SG that *ec* go.IMPF.3SG travel.INF
'Pedro said that he was going to travel'

c. Antigamente __ punha mesa para tomar lanche (Duarte 2000)
formerly *ec* put.IMPF.3SG table for take.INF snack
'Formerly one used to lay the table for snacks'

One issue that these examples illustrate is that the claim that BP has lost the null subject parameter does not mean there are no null subjects at all in BP. Rather, it means that null subjects are less frequent than lexical subjects; and, more importantly, it means that very strict conditions apply in order for them to be properly interpreted. If we examine all the subjects represented by the gaps in (10) above, we will see that in (a), the only interpretation for the subject of *quer* 'wants' is as co-referent with the subject of *sabe* 'knows' in the main clause (*ninguém* 'nobody'); in (b), the subject of *viajar* 'to travel' can only be interpreted as co-referent with the subject of *disse* 'said' in the main clause (*O Pedro*); and in (c), the subject of *punha* 'used to lay' can only receive an indeterminate interpretation.[3] The particular conditions involved in the interpretation of the empty subjects in these examples illustrate, broadly, what generative theory refers to as "restrictions" on the licensing of null subjects. And, because "subject" is a central grammatical category (among others, it is the position reserved for the most prominent argument of the verb), the extent to which a language licenses/limits the expression of this category as null (if at all) is central to its whole organization, and has consequences for several, seemingly unrelated issues. Also, the way a language differs in this respect is central to how it can be differentiated from other languages (which, again very roughly, is what the term "parameter" means in this context).

Generative research on BP, in sum, claims that the loss of the *null-subject parameter* provoked a complete grammatical reorganization, and is at the root of most of the other important syntactic properties of the language, of which we will discuss, below, the order of constituents (3.2), and the conditions for argument movement (3.2.1, cleft-constructions and, 4, strategies for relativization).

As for complement expression, BP syntax displays two particular phenomena. The first one involves the possibility of omitting the direct complement in cases where it provides information which can be recovered from the context, as illustrated in (11) below (example from Galves 1989):

(11) No tempo de calor a gente colhe as
in-the time of heat the people take.PRES.3SG the

maçãs e guarda __ no porão para
apples and keep.PRES.3SG *ec* in-the cellar to

3 Notice that none of this is true for other varieties of Portuguese: both in Classical and European Portuguese, all the "gaps" in example (10) would be interpretable as linked to a definite referent mentioned in the prior discourse.

comer no inverno
eat.INF in-the winter
'In summer we take the apples and keep them in the cellar to eat them in the winter'

The second phenomenon, which commonly occurs in informal varieties, involves the frequent use of a subject pronoun in a context where atonic clitics would be expected in formal registers, as shown in (12) below:

(12) Eu emprestei *ele* ao colega ~ Eu emprestei-*o*
I borrow.PST.1SG it.NOM to-the colleague I borrow.PST.1SG-it.ACC
ao colega
to-the colleague
'I lent it to the colleague'

The possibility of omitting direct complements and the use of subject-pronouns in object positions are correlated to the fact that object clitic pronouns are falling into disuse in BP. In cases such as (11) above, the insertion of a third person clitic pronoun as the argument of *guarda* 'keep' (forming *e **as** guarda* 'and keep them') is not common in BP; and in cases such as (12), the use of a third person clitic *o* 'it' instead of the nominative *ele* would only be common in the written register. In both cases, the use of clitics would sound very formal, whereas in European Portuguese they are part of the general vernacular. For a detailed account of the diachronic process involved in this change, see ↗6 Historical syntax.

In formalist research, "null" objects are also conceptualized as empty grammatical categories, and their use in BP has often been studied in relation to the major grammatical change pertaining to the expression of subjects. In fact, both the possibility of null complements and the use of subject-pronouns in complement position may be seen as consequences of the reorganization of the pronominal system associated with the loss of the null-subject parameter (see Galves 1989; Kato 1993a; Cyrino 1996).

3.2 Order of constituents

As to the positioning of sentence constituents, the different ordering patterns in BP have been extensively described as sensitive to pragmatic conditions. This is a fundamental element in understanding BP syntax; and, beyond the points to be dealt with in this section (topicalization, focalization—including cleft constructions, and voice), more detailed discussions can be found in Castilho (2002), Ilari (2003), and Neves (2003), respectively volumes I, II and VII of *Gramática do Português Falado*.

These and other studies have shown that the most frequent patterns in BP are SV(O) and VS(O) (Pezatti 2014; Camacho/Pezatti 1997; Pezatti 1993). These patterns show a co-relation, in a complementary distribution, to verb types: SV(O) co-relates to tran-

sitive, intransitive and copula verbs (*eles preferem linguística* 'they prefer linguistics'; *o rapaz trabalha duro* 'the young man works hard'; *a regra é flexível* 'the rule is flexible'), whereas VS(O) covers ergative constructions with existential, presentational verbs (*existem pessoas em condições privilegiadas* 'there are people in a privileged condition'; *veio um senhor de meia idade* 'there came a middle-aged man').

Functionalist studies have noted that the interaction between the constituents' informational status and the pragmatic functions of topic and focus help in understanding the choice of one or the other pattern. Camacho/Pezatti (1997) propose a general order pattern, represented by P1 (V) S (V) O (V), which accounts for alternative verb (V) positions and defines P1 as the initial position to be filled by *wh*-type grammatical constituents, relative pronouns and subordinate conjunctions, or, in their absence, by constituents which take on topic or focus functions.[4] The abbreviations S and O refer to subject and complement, respectively.

The P1 (V) S (V) O (V) scheme can be used to describe the most frequent order patterns in BP systematically. In (13), the grammatical function of subject and the pragmatic function of topic overlap in P1 (*a produção* 'the production'). In this case, the constituent's topical status depends on a context in which (13) answers the question: *Has the production increased?* Conversely, (14) to (16) include cases with no pragmatic motivation to place the subject in P1, since it does not have the function of topic. In the ergative structure in (14), P1 remains empty and the subject is post-verbal. In (15), P1 is filled by the *wh*-word *onde*, inherently focal, a property which is reinforced by the cleft *é que*. In (16), P1 is filled by the topic (*cem mil cruzados* 'one hundred thousand cruzados') and the VS structure, as a block, provides focal information which is usually new (examples from Camacho/Pezatti 1997):

(13) A produção cresceu muito (P1/S V O)
the production increase.PERF.3SG much
'The production has increased much'

(14) Expirou o prazo (__ V S)
expire.PERF.3SG the deadline
'The deadline has expired'

(15) Onde é que estão os economistas? (P1 V S)
where É QUE be.PRES.3PL the economists
'Where are the economists?'

4 The *topic* function is assigned to the constituent comprising the entity about which a predication is made, and the *focus* function to the constituent which holds the most important or outstanding information, considering the communicative situation (Dik 1989).

(16) Cem mil cruzados faturou nossa barraca. (Votre/Naro 1989)
hundred thousand cruzados earn.PST.3SG our stall (P1 V S)
'Our stall earned 100.000 cruzados'

In generative syntax research, the order of constituents in BP is taken as one of the most significant features in the study of its grammar, and is seen as closely related to the loss of the null-subject parameter. This can be explained by the general observation that languages which show the frequent use of null subjects also often have rich verb-agreement morphology, and relative freedom in the ordering of constituents; and that, in historical processes, the three properties tend to be lost together. This is the case in Portuguese, where BP has lost, in the same process, null subject, rich subject-verb agreement morphology, and freedom in word order.

Also, for formalist studies, different orders of constituents are taken as related to different informational properties; therefore, the variation between different languages in this regard will express a difference in the relation between syntax and information structure in each grammar (and also other linguistic levels, notably phonology). In this context, the interaction between discourse and syntax has been a central factor of the research into the properties of the order of constituents in BP, particularly since Negrão (1999).

In the following, we discuss two phenomena which interact in a particularly direct manner with the order: topicalization and focalization strategies and voice constructions.

3.2.1 Topicalization and Focalization strategies

There are two key pragmatic functions involved in the linguistic organization of information: *topicality*, which characterizes the starting-point, or topic of the messages; and *focality*, which characterizes the most salient aspects of the information relating to a topic. The pragmatic functions *topic* and *focus* are indicated by phrasal constituents treated as topical or focal, respectively (footnote 4). BP, as with other languages, presents specific syntactic strategies that allow for the unambiguous marking of topical and focal phrasal constituents, some of which we will summarize here.

3.2.1.1 Cleft constructions represent a syntactic focus mechanism which rearranges the sentence constituents in a type of equational structure which establishes a relationship of identification (Halliday 1985), where the identifier function, mapped in the focal constituent, creates a sense of exclusivity which triggers contrastive readings. Therefore, cleft sentences are highly contrastive. BP presents at least five cleft variants, illustrated below, whose structural configurations and conditions of use differ

slightly (Braga 1991; Longhin 1999), as shown in (17) to (21) (examples taken from Longhin 1999):[5]

(17) Regular cleft constructions (CLIV):
A agricultura de ciclo anual não vive e é essa
the agriculture of cycle yearly not live.PRES.3SG and CLIV this

agricultura que alimenta o homem.
agriculture CLIV feed.PRES.3SG the man
'The yearly agricultural season does not survive and it is this agriculture that feeds mankind.'

(18) Constructions with "é que" (É QUE):
A física também tinha estagnado depois de um grande avanço,
the physics also stagnate.PST.PERF.3SG after of a big advance

depois da física nuclear, e não estava produzindo coisa nova.
after of-the physics nuclear and not produce.PST.PROG.3SG thing new

O novo vinha da genética. De fato, a genética é que
the new come.IMPERF.3SG of-the genetics of fact the genetics É QUE

produziu uma série de transformações.
produce.PST.3SG a series of transformations
'Physics had also stagnated after a great advance, after nuclear physics, and was not producing anything new. What was new came from genetics. In fact, it was genetics that produced a series of transformations.'

(19) Constructions with que (QUE):
Não sou feia não, viu? ***Você*** **<u>que</u>** **é** **feia**.
not be.PRES.1SG ugly no see.PST.3SG? You QUE be.PRES.3SG ugly
'I am not ugly, do you hear? You are ugly.'

(20) "To be" focus (SF):
Põe lá na caixa. Não, **vou pôr** **<u>é</u>** ***aqui***.
put.IMP.3SG there in-the box no put.FUT.1SG SF here.
'Put it there in the box. No, I will put it here.'

5 In these examples, we have used bold to mark the whole cleft sentence, underline to indicate the formal cleft marks, and bold-italic for the focal constituent.

(21) Pseudo-clefts (PC):

É	que	aqui	falta	comida,	o	pessoal	padece	
be.PRES.3SG	that	here	lack.PRES.3SG	food	the	people	suffer.PRES.3SG	
fome.	Um	país	tão	grande	e	organizado	tão	mal.
hunger	a	country	so	big	and	organized	so	bad
O que	nós	somos	é	um	país	faminto.		
PC	we	be.PRES.3SG	PC	a	country	hungry		

'It happens that here we don't have enough food, people suffer famine. Such a large country and so poorly organized. What we are is a hungry country.'

Functionalist-oriented research (Braga 1991; Longhin 1999) has suggested that a strong co-relation exists between the type of cleft, the order of the constituents, and the informational status of the element in focus. "Regular" clefts, cleft constructions with *é que*, and cleft constructions with *que* (CLIVs, É QUEs and QUEs in 17–19 above) are more likely to put into focus information which is already known in the communicative situation, whereas constructions with 'to be' focus and pseudo-clefts (PCs and SFs in (20) and (21) above), by turn, are preferred for focalizing new information. The position of the focal constituents in cleft sentences can be explained, at least to some extent, by the principle of information distribution, according to which new elements tend to follow given elements. Thus, since the focus of PCs and SFs tends to codify new information, it will most naturally appear to the right, and since the focus of CLIVs, É QUEs and QUE tends to codify known information, it will most naturally appear to the left.

In formalist syntax, the properties of cleft constructions in BP have been investigated by Kato et al. (1996), Modesto (2001), among others. This approach views cleft constructions as complex sentences formed by a clause headed by an auxiliary verb merged with a second clause in which one of the arguments is co-indexed (i.e., they share the same referent) with an argument of the first clause; this results in a construction in which the two clauses form one and the same logical proposition, and which present specificational readings such as contrastivity, exclusivity, and exhaustivity. Yet such constructions can be analyzed differently by different researchers; in the examples below, for instance, *que* is not treated as an argument of the verbs in the second clauses, but rather as a phrase-marker, or *complementizer* (marked COMP in the examples; the index (i) represents co-indexing of constituents (i.e., sharing of referents), and represents a gap in the structure:

(22) é **<u>essa agricultura</u>$_{(i)}$** que-COMP $\_{(i)}$ alimenta o homem.

(23) De fato, **<u>a genética</u>$_{(i)}$** é que COMP $\_{(i)}$ produziu uma série de transformações.

One of the main points of interest surrounding cleft-constructions from a formalist perspective is understanding the mechanisms by which the movement of the argument of one verb to the domain of another verb, within the same sentence domain, may trigger such specificity readings. Cleft-constructions in BP are particularly interesting in this line of research, since they reveal important properties of the relation between the conditions on predication and the restrictions on the movements of argumental and non-argumental constituents.

3.2.1.2 Topicalization constructions. As briefly mentioned above, the pragmatic property of *topicality*, which characterized the starting-point of the messages, may be expressed by different structural strategies through which a language indicates the function "topic" in a phrasal constituent. Among those strategies, the order of constituents is, in itself, strongly related to topicality. This may be seen in the following examples, where the left-most phrasal constituents, highlighted in each sentence, may be analyzed as *topics* from different perspectives, as we will subsequently describe in more detail:

(24) **Essa competência** ela é de natureza mental (Pontes 1987, 19)
this competence she be.PRES.3SG of nature mental
'This competence, it is of a mental nature'

(25) **A Rosa** eu falei com ela ontem (Pontes 1987, 14)
the Rosa I speak.PST.1SG with her yesterday
'Rosa, I spoke to her yesterday'

(26) **Essa torneira aí** não sai água? (Pontes 1987, 35)
this faucet there not come-out.PRES.3SG water
'This faucet, does no water come out of it?'

(27) **Esse rádio** estragou o ponteiro (Pontes 1987, 31)
this radio-set spoil.PST.3SG the marker
'This radio set has a broken marker'

(28) **A lanterna**, já comprou as pilhas? (Pontes 1987, 14)
the flashlight already buy.PST.3SG the batteries
'The flashlight, did you buy the batteries yet?'

(29) **Eu agora**, acabou desculpa de concurso, né? (Pontes 1987, 13)
I now finish.PST.3SG excuse of exam, no
'For me there is no more the excuse of exams'

(30) **Eu**, café eu gosto tanto sem açúcar como com (Pontes 1987, 30)
I coffee I like.PRES.1SG much without sugar as with
'I myself like coffee both with and without sugar'

The classical examples from Pontes (1987), reproduced in (24) to (30) above, show salient distinctions in their structural configuration and discoursive functioning. As regards their structural configurations, we can identify:

(i) Constructions in which topics and subjects are co-referential (24)
(ii) Constructions in which the topics and (part of) the indirect objects are co-referential (25)
(iii) Constructions in which the co-reference is indirect or associative, grounded in *frame*s (26–28)
(iv) Constructions that involve more than one topic (30). In this example *café* 'coffee' is of course the topic (the element of which something is predicated, i.e., which one 'likes with or without sugar'); but this is joined by another topic, *eu* 'I', which takes over the property of topicality by reason of contrastivity (i.e., 'I' versus 'others', or other participants in the context of enunciation).

Functionalist research (Braga 1987, among others) considers the aspects of form and meaning in topic constructions together, with the main goal of showing the discoursive functions of those structures in actual communicative situations. Braga claims that topic constructions like (31) below, with the direct object dislocated to the left, have the essential functions of casting out one of the elements from a group, referring back to elements mentioned in the previous discourse, or setting contrast relations. Constructions like (32), in turn, function essentially as strategies to introduce and reintroduce topics in the discourse (examples from Braga 1987):

(31) Assim, **segredo**, só falo pra ela
so secret only speak.PRES.1SG to her
'Secrets, I only tell her'

(32) **O Éder**, no outro jogo, ele não foi
the Éder in-the other game, he not go.PST.3SG
'Éder, in the other game he didn't show up'

In generative syntax, *topicality* would be seen as a (stable and universal) pragmatic function, and the operations of *topicalization* as one of the points of interface between pragmatics and syntax. Topicalization constructions operate within the syntactic

component, where the sentence may be re-structured—re-ordered, for instance—so as to satisfy *pragmatic* (or *discoursive*) requirements. Grammars will vary regarding the form of this interaction, because the options open for restructuring depend on the general syntactic properties of the language itself: specifically syntactic requirements (for instance, the limits and rules for the expression of arguments) will combine with pragmatic requirements and result in different outputs, different *topicalization constructions*, in each language. In this sense, the topicalization constructions characteristic of BP are seen once more as illustrative of the restriction on the null subject's parameter, since, for strictly syntactic reasons, this grammar presents a strong preference for SV order. The coincidence of this particular syntactic requirement and the universal pragmatic requirement of marking topicality would explain, for instance, the characteristic topicalization construction in which there is a topic and a lexical (pronominal or nominal) pre-verbal subject, as in examples (33) and (34) above, analyzed below:

(33) [TOPIC **Essa** **competência**$_{(i)}$] [**ela**$_{(i)}$ é de natureza mental]
this competence she be.PRES.3SG of nature mental
'This competence, it is of a mental nature'

(34) [TOPIC **A** **Rosa**$_{(i)}$] [**eu** falei com ela$_{(i)}$ ontem]
the Rosa I speak.PST.1SG with her yesterday
'Rosa, I spoke to her yesterday'

Note, in this analysis, that the topic *Essa competência* shares its referent with the subject of the clause, *ela*, in (33), as indicated by the index (i); the same applies to *A Rosa* and the complement *ela* in (34). An interesting avenue of research here would be to ask what are the conditions for the sharing of referential interpretation between those elements, when one of them is beyond the predication domains of the verb (*Essa competência*; *A Rosa*)? Topicalization, in this sense, is a promising field of research in generative syntax, as it points to the challenges of exploring an important grammatical interface that looks beyond the sentence. The characteristic constructions of BP here make it a good candidate for further research in this regard.

3.2.2 Voice

Voice options represent alternative wordings which reveal different perspectives or points of view for the same content, derived from a change in the informative weight of the argumental constituents of the sentence, or possibly the omission of one of them. In these terms, functionalist literature has approached voice as primarily a pragmatic-discoursive mechanism (Camacho 2002; Ilari 2014).

Different syntactic order patterns mark functional voice options, whereby pieces of information can be highlighted, put into the background, or even hidden, as a means of creating particular meanings. In BP, the motivation for placing the agent in the final position often lies in the need to mark it as *new* in the communicative situation; on the other hand, the reason for omitting the agent, which is common for certain kinds of passive sentences (for example, *eu nunca fui assaltado* 'I have never been robbed') is the retrievability of the information, the impossibility of identifying the agent, or even its low informational relevance, as judged by the user.

Among the various voice alternatives, in BP we highlight three passive voice constructions which share the fact that the agent is hidden, even though they are based on different structural settings. These are passive constructions formed by the periphrasis *ser* 'be' *+ past participle*, impersonal constructions with the passive clitic pronoun *se*, and constructions with *ter* 'have' *+ past participle*, as shown in (35), (36) and (37), respectively:

(35) O condomínio foi construído recentemente (pelo empreendedor)
the condominium be.PST.3SG build.PTCP recently (by-the entrepreneur)
'The condominium was built recently (by the entrepreneur)'

(36) Aluga(m)-se ternos
rent.PRES-PASS suits
'Suits for rent'

(37) Diretor da empresa tem telefone grampeado (pela polícia federal)
director of-the company have.PRES.3SG telephone wiretap.PTCP (by-the police federal)
'Company director has his telephone wiretapped (by the federal police)'

In (35) and (37), the grammatical subject is mapped in the patient, which acts as a topic, that is, the starting point of the sentence. In this case, omitting the agent is optional. In (36), the pronoun *se* does not display reflexivity and reciprocity properties, playing the role of making the agent indeterminate. In this case, the patient, usually an inanimate entity, is mapped in the post-verbal grammatical subject, the single argument. The difference in the perspective of the two passive sentences is clear: the choice between a construction with an auxiliary verb and an impersonal one is strongly related to the fact that the first allows an agent noun phrase to be inserted and displays a topic, features not available for the impersonal construction, which comes closer to being a subjectless sentence.

In formalist accounts, the term "voice" is used less than in functionalist theory. This may be related to the fact that for formalists, the notion of the grammatical subject being "mapped" in different thematic roles—agent, patient—is not a valid concept.

Rather, the reverse is the case: arguments with different thematic roles will be projected, or not, as subjects, depending on each construction. Therefore, "passive", "active" and "middle" are not seen as properties of the interaction between the pragmatic, semantic and syntactic levels: formalist will refer to "active" or "passive" *constructions* as different syntactic realizations of the same argument structure. It might be interesting here for us to analyze the same sentences again, from a formalist perspective:

(38) O condomínio foi construído recentemente (pelo empreendedor)

(39) Aluga(m)-se ternos

Let us focus on sentence (38) above. This is the passive construction of the argument structure [*construir*: *condomínio*, *empreendedor*], and it would be analyzed, formally, as an alternative predication to this argument structure, triggered by the formal trace of "passive" in the auxiliary *ser* (*to be*, in the past tense, *foi*), in contrast to the active predication, which is constructed in the absence of this trigger. In the sentence, the subject is [*o condomínio*]. This brings no special provision to the analysis, since from this perspective there is no direct mapping of thematic roles and the property "subject": the subject is the argument that occupies a prominent position, and that establishes a special relation with the verb (which can, though not obligatorily, present itself as an "overt agreement" or morphological agreement). The subject may have different thematic roles depending on the structure; in passive constructions, it will be the patient. This becomes interesting when we turn to sentence (39) above. In a functionalist analysis, this sentence is analyzed in the same way, whether the verb presents agreement with the subject or not: *aluga(m)-se*. However, in a formalist analysis, the presence or absence of agreement makes all the difference. If the sentence is like (40) below (with evidence of agreement), then *ternos* is the subject. If the sentence is like (41) below (with no evidence of agreement), there is an alternative analysis, in which *ternos* is not the subject of this sentence, but, rather, is the object; and the subject of the sentence (surprisingly) is the pronoun *se*:

(40) Alugam-se ternos-*Subject*

(41) Aluga-se-*Subject* ternos-*Object*

In this analysis, example (41), crucially, is not a passive construction. Rather, it is an active construction in which the subject has the particular (referential) property of indetermination. In formalist studies on BP there is considerable interest in the properties of BP grammar in terms of the problem of indeterminate subjects and non-active subjects, as can be seen in Duarte/Kato/Barbosa (2001) and Negrão/Viotti (2008; 2015), among others.

3.3 Complex predicates

In this session, we discuss constructions in which the verb is not the sole semantic base for the construction of the clause, that is, where the verb is not the only predicator (as has been the case in what we have described thus far). Instead, the verb carries syntactic-semantic functions, as we will now see regarding *light verbs* (3.3.1), and in *auxiliary verbs* (3.3.2).

3.3.1 Light verbs

Constructions with a light verb are composed of two parts, *verb + noun phrase*. The light verb (usually *dar* 'give', *levar* 'take', *tomar* 'have'/'take', *fazer* 'do') carries the grammatical categories (mood, tense, number and person), but it is not the only decisive element for the argument structure: this is also enacted by the noun phrase, which plays a significant role in the predication. This can be seen in (42), where the non-referential NP *uma análise* 'an analysis', rather than playing a semantic role as a participant in the event expresses by *fazer* 'to make', in fact functions as a predicator; it co-determines, with the verb 'to make', the arguments *João* 'John' and *a situação* 'the situation'. The proposition *João fez uma análise da situação*, while literally translated as 'João made an analysis of the situation', in fact means 'João analized the situation', with *fez uma análise* 'made an analysis' as a "periphrase" of *analisar*, 'to analyze':

(42) João **fez** **uma** **análise** da situação
John make.PST.3SG a analysis of-the situation
'John made an analysis of the situation'

In BP, constructions with light verbs show different levels of idiomaticity, in which highly formulaic constructs, such as (43), coexist with morphosyntactically flexible structures, such as [*dar uma X-*$_{da}$], which can be filled in various manners, as in (44) and (45).

(43) A seleção **deu** **um** **banho** de técnica e ousadia (= excelência)
the national team give.PST.3SG a bath of skill and courage
'The national team gave a master performance of skill and courage'

(44) Quando tinha tempo, **dava** **uma** **limpada** na casa
when have.IMPERF.1SG time give.IMPERF.1SG a clean-up in-the house
'When I had time, I would give my house a clean'

(45) Eu **dei** **uma lida** no artigo de linguística
I give.PST.3SG a reading on-the article of linguistics
'I had a look at the linguistics paper'

Examples (44) and (45) make clear that constructions with light verbs and their counterparts with full verbs (*limpava a casa* 'cleaned the house'; *leu o artigo* 'read the article') do not create the same meaning, despite being legitimate semantic functional options. Apart from the mitigating effect of (42) and (43), functionalist-based studies (Neves 2002; 2006; Ilari 2014) have supported the notion of syntactic versatility (e.g., *tomar a decisão final*/*decidir* 'to take the final decision/to decide'), semantic precision (e.g., *tomar a decisão final* vs. *decidir finalmente* 'to take the final decision/to finally decide') and appropriateness of register (e.g., *fazer xixi* 'to pee') which make constructions with light verbs unique.

Formalist studies of BP have also paid special attention to constructions with light verbs (see Scher 2006, among others), in that they reveal interesting and unique characteristics in the properties of predication in this grammar. As we noted above, one approach to constructions with light verbs is to see the light verb as supporting grammatical features but having no argument structure. On this view, the clause structure in such cases would be similar to constructions with auxiliary verbs (i.e., complex predicates), as illustrated in (46) below; a different analysis, however, would be to say that the light verbs still carry predicative properties (i.e., they are still argumental verbs), as illustrated in (47) below:

(46) Eu$_{(i)}$ **dei** uma __$_{(i)}$ **lida** no artigo de linguística

(47) Eu-*Subject* **dei** uma lida-*Object* no artigo de linguística-*Oblique*

What would single out those constructions, in this case, would be the particular argument structure of a class of verbs (and the potential for, and constraints on, this change over time). Constructions with light verbs, therefore, help us to understand better the interaction between argument structure and phrase construction (i.e., the projection of argument structure). Note that, if we follow the concept of phrase construction in this theory, the sentence is the projection of the argument structure of the verb, and this, in turn, is "a given" (i.e., it is pre-established). This does not mean, however, that there isn't any interest in formalist syntax for argument structure; on the contrary, many studies have been dedicated to this, and such work relies heavily on the phenomenon of light verbs.

3.3.2 Auxiliary verbs and other verbal periphrases

A characteristic of BP syntax, resulting from the general tendency for analysis which permeated the development of Romance languages, is the great variety of V1 V2-type verbal periphrases, in which V1 is the auxiliary verb centering the grammatical information, and V2 is the full verb and defines the semantic conditions in regard to other sentence elements. Together, V1 and V2 form a construction which equates to a simple verb. In the set of verbal periphrases, we observe those formed by gerunds (*estar + -ndo*, *ir + -ndo*, *acabar + -ndo*, *vir + -ndo*, *ficar + -ndo*, *continuar + -ndo*), by infinitives (*ter que + -r*, *dever + -r*, *saber + -r*, *conseguir + -r*, *deixar de + -r*, *acabar de + -r*, *passar a + -r*, *terminar de + -r*) and by past participles (*ser + -do*, *ter + -do*, *estar + -do*, *ficar + -do*).

Research on periphrases in BP (Neves 2006; Ilari 2014, among others) has shown that, within this set, genuine auxiliary verbs coexist with others which relate only partially to the auxiliaries. This fluid categorization can be explained in light of the grammaticalization processes which have affected full verbs, pushing them towards a state of auxiliariness. Since grammaticalization processes occur over the course of centuries and might indeed be ongoing, it is reasonable that not all traces of the new category will be assimilated, which explains the fluidity.

From this perspective, some notable studies on this aspect of BP syntax (Ilari 2014) have defined criteria or tests to assess the level of auxiliariness (e.g., subject sharing, scope of negation, presence of intervening material between V1 and V2, semantic emptying of V1 etc.), the implementation of which allows us to observe that most periphrases pass the auxiliariness test, while others fail on one or another criterion, revealing an as yet incomplete course of auxiliarization and thereby capturing the gradual nature of verb categories.

Also in the domain of verbal periphrases, BP uses a complex construction, innovative in the diachrony of Portuguese, which is common in more informal spoken and written registers, described according to its aspectual value, or indeed based on its value of overturned expectation and emphasis (Tavares 2008; Rodrigues 2009). From a morphosyntactic point of view, it is a periphrasis formed by a minimal and invariable sequence of two verbs, V1 and V2, which share subject and the verb inflections. The verbs which can fill position V1 form a restricted group, in which *ir* 'go', *chegar* 'arrive', *pegar* 'take', *vir* 'come' and *virar* 'become' are the most recurrent. The verbs which fill V2, in turn, are a relatively open class. V1 and V2 appear connected by the conjunction *e* or can be juxtaposed, corresponding to the types (48) and (49) (examples from Tavares 2008; Rodrigues 2009, respectively):

(48) No cinema, não via as letras, minha filha. E o povo
in-the cinema not see.IMPERF.3SG the letters my daughter and the folks

ria, e eu ria. O povo ficava sério,
laugh.IMPERF.3SG and I laugh.IMPERF.1SG the folks get. IMPERF.3SG serious

e eu ficava séria. Eu só via a imagem. Aí eu
and I get. IMPERF.3SG serious I only see.IMPERF.1SG the image. Then I

peguei e pedi pro meu noivo trazer né? A fita de vídeo pra mim ver.
ask.PST.1SG to-the my fiancée bring.INF, no the videotape for me see.INF
'In the cinema I could not see the letters, my child. And the folks laughed, and I laughed. The folks got serious, I got serious. I could only see the image. Then I asked my fiancée to bring it, right? The videotape for me to see.'

(49) Ele atravessou na frente do carro, não é? O carro foi,
he cross.PST.3SG in front of-the car no be.PRS.3SG The car go.PST.3SG

jogou ele para o alto, caiu na calçada.
throw.PST.3SG him into the high fall.PST.3SG on-the sidewalk
'He crossed in front of the car, right? The car came and threw him into the air, fell on the sidewalk.'

The periphrases *V1 (e) V2* formally and pragmatically approximate to paratactic constructions, which leads Rodrigues (2009) to develop a hypothesis about an existing relatedness in which the periphrases would have been grammaticalized based on parataxis. According to that author, just as in parataxis, in the periphrases *V1 (e) V2* the second member of the construction is in focus. The process of change consisted of the decategorization of V1 and, subsequently, full reanalysis of the construction. In the context of periphrasis, V1 underwent syntactic-semantic changes, such as lexical meaning opacification and loss of the property to select arguments, which led its full verb status to be weakened and a new discoursive-pragmatic function to be obtained, namely that of dramatizing or emphasizing the events expressed in V2. Thus, in the periphrases *V1 (e) V2*, V1 creates the conditions to add V2, putting it into a prominent position, where V2 adds new and/or contrastive, usually unexpected or surprising information in the communicative situation.

4 Relativization strategies

Since the groundbreaking work of Tarallo (1983), BP relativization strategies have been one of the most researched aspects of the grammar from any theoretical outlook. The main generalization that can be made here is that BP shows a preference for constructions such as (51) and (52), rather than (50), that is, a preference for "chopping"

relatives or "resumptive" relatives, rather than "standard" relatives. Some classic examples from Tarallo (1988, 141–142) follow:

(50) Standard relative:
E um deles foi esse fulano aí, **com quem eu nunca**
and one of-them be.PST.3SG this chap there, with whom I never
tive aula
have.PST.1SG class
'And one of them was that chap with whom I never had class'

(51) Resumptive relative:
E um deles foi esse fulano aí, **que eu nunca tive**
and one of-them be.PST.3SG this chap there, that-REL I never have.PST.1SG
aula com ele
class with him
'And one of them was that chap who I never had class with him'

(52) Chopping relative:
E um deles foi esse fulano aí, **que eu**
and one of-them be.PST.3SG this chap there, that-REL I
nunca tive aula
never have.PST.1SG class
'And one of them was that chap with whom I never had class'

According to functionalist syntax, BP displays singularities regarding relativization strategies, having a typology in which relative constructions are not free variants, but are choices governed by decisions in the scope of semantic and pragmatic components, which amounts to saying that relativization strategies are linked to different communicative purposes (Camacho 2013; 2017). Communicative intentions along with the fact of change explain, to a great extent, the variable morphosyntactic encoding of relative clauses, which predicts relative constructions with and without a preceding noun, as in (53a–b); relative clauses embedded or not in an NP, as in (54a–b); relative clauses headed by a relative pronoun and relative clauses headed by conjunctions, as in (55a–b).

(53) a. **a recepção** que foi preparada pelos alunos
the reception that.REL be.PST.3SG prepare.PTCP by-the students
superou as expectativas
exceed.PST.3SG the expectations
'The reception which was prepared by the students exceeded expectations'

b. **quem já deu** não tem para dar
who-REL already give.PST.3SG not have.PRES.3SG to give.INF
'They who have already given have nothing to give'

(54) a. O médico **que fez o parto** deu entrevista
the doctor that.REL do.PST.3SG the childbirth give.PST.3SG interview
'The doctor who delivered the child gave an interview'

b. O João, **que é um excelente aluno**, atingiu
the John who.REL be.PRES.3SG an excellent student reach.PST.3SG
todas as metas
all the goals
'John, who is an excellent student, reached all the goals'

(55) a. O time **que** não treina não tem sucesso
the team that.REL not practice.PRES.3SG not have.PRES.3SG success
'A team which does not practice does not succeed'

b. Não conheço o rapaz **que** o João emprestou o livro
not know.PRES.1SG the guy that.REL the John borrow.PST.3SG the book
'I don't know the guy who John borrowed the book [from]'

In the context of the possibilities of relativization, functional research has focused on aspects related to the generalization process which has affected the relative pronoun *que* (roughly equivalent to the English relative 'that'), especially in oral enunciations, narrowing the set of relative pronouns (*quem* 'who', *qual* 'which', *cujo* 'whose', *onde* 'where', and *quanto* 'how much'), given that, due to changes that have occurred, *que* has expanded contextually, taking the place of other relative pronouns and therefore being established as a universal relative pronoun (Castilho 2010).

As to the typology of relative clauses, some have the purpose of *identifying* or *restricting* a subset within a greater reference set, by fulfilling a condition for the noun core. The group of restrictive clauses includes standard constructions, more frequent in the formal written register, in which the participating relative pronouns show causal specifications, as in (56) to (59).

(56) Despachei uma encomenda para meu tio **que mora** (*Subject*)
dispatch.PST.1SG a parcel to my uncle that-REL live.PRES.3SG
na capital
in-the capital
'I sent a parcel to my uncle who lives in the capital'

(57) A cidade **em que** **nasci** fica na região noroeste
the city in that-REL be-born.PST.1SG be.PRES.3SG in-the region northwest
'The city where I was born lies in the Northwest' (*Oblique*)

(58) Comi o bolo **que** **eu** **mesma** **fiz** (*Direct complement*)
eat.PST.1SG the cake that-REL I myself make.PST.1SG
'I ate the cake which I baked myself'

(59) Este é o professor **a** **quem** **me** **refiro** (*Indirect complement*)
this be.PRES.3SG the professor to whom-REL I refer.PRES.1SG
'This is the professor whom I refer to'

Turning to restrictive clauses, BP displays two particular patterns, both strongly related to the depronominalization of the relative pronoun, which involves loss of phoricity and a reinforced conjunction status, decisive transformations which make it come closer to being a complementizer. In one of these patterns, seen in (60) below, the relative clause carries a personal pronoun, usually joined by a preposition, which ensures co-reference with the preceding NP. These are the so-called "copying relative clauses" (see Tarallo 1983; Camacho 2013, among others), constructions which are stigmatized in the context of more formal language use.

(60) A professora **que** **você** **gosta** **<u>dela</u>** vai se aposentar
the professor that-REL you like.PRES.3SG of-her retire.FUT.3SG
'The professor you like will retire'

The second specific pattern covers restrictive relative clauses whose main characteristic is the lack of a co-referential anaphoric element. They are the so-called "chopping relative clauses", as in (61), which are recurrent in spoken and written genres in BP.

(61) A professora **que** **você** **gosta** vai se aposentar
the professor that-REL you like.PRES.3SG retire.FUT.3SG
'The professor who you like will retire'

Another kind of relative clause, different from the restrictive one, is motivated by the aim of adding supplementary information, an apposition, to the referent of the preceding NP, which has been defined in the usual way. These are called *appositive relative clauses*. In this case, the constructions are realized in a non-embedded mode of syntactic composition, in which the main clause and the relative clause have different illocutionary forces, as in (52b) mentioned above.

It should be added that BP also has relative clauses with the specific ability of enabling interpretations typical of circumstantial clauses, as in (62) to (64), which can be read in terms of cause, contrast and condition, respectively. In this case, the circum-

stantial readings are strongly pragmatic, and depend on linguistic and pragmatic-cognitive contextual factors.

(62) Meu irmão, **que morou na Itália**, conhece bem o percurso
my brother that-REL live.PST.3SG in-the Italy know.PRES.3SG well the route
'My brother, who lived in Italy, knows the route well'

(63) O atleta, **que acumulou glórias no futebol**, morreu esquecido
the athlete that-REL accumulate glories in-the soccer die.PST.3SG forgotten
'The athlete who received much glory in soccer died in obscurity'

(64) Seriam selecionados todos que tivessem
be.COND.3PL select.PTCP all that-REL have.SUBJ.PST.3PL
experiência internacional
experience international
'All who had international experience would be selected'

For generative syntax, the strategies of relativization in BP have long been an intense field of research, indeed, ever since Tarallo's (1988) influential study and its notable repercussions, as Kato (1993b) has shown. From this perspective there was particular interest in what the relative constructions in BP might reveal about the conditions of the movement of constituents to different parts of the sentence structure. For some researchers here, the contrast of standard and non-standard relatives in BP has been linked to general restrictions on the movement of argumental constituents in BP (as already mentioned for Cleft-constructions), with the options involving less movement being preferred. A summarized typology of one of the possible analyses in the generative literature is presented below, based on sentences already shown and glossed above:

(65) Indirect Complement relatives:

a. ...esse fulano aí$_{(i)}$, [com quem$_{(i)}$ eu nunca tive aula] (*Standard)*
b. ...esse fulano aí$_{(i)}$, [que-COMP eu nunca tive aula com ele$_{(i)}$] (*Copying)*
c. ...esse fulano aí$_{(i)}$, [que-COMP eu nunca tive aula __$_{(i)}$] (*Chopping)*

d. Esse é [o professor ($_{i}$) [a quem$_{(i)}$ me refiro__$_{(i)}$] (*Standard*)
e. [A professora$_{(i)}$ [que-COMP você gosta dela$_{(i)}$]] vai se aposentar (*Copying*)
g. Não conheço [o rapaz$_{(i)}$ [que-COMP o João emprestou o livro __$_{(i)}$] (*Chopping*)

In this analysis, *que* in "copying" and "chopping" relatives is not a relative pronoun (as *quem* is), but is a complementizer (similar to the analysis for clef-constructions, above). Note, however, that this analysis applies most clearly to indirect complement relatives, such as the ones shown above; for direct complement relatives and subject relatives, there is an analytical ambiguity between the standard and the

chopping structure; this can be seen in examples (58) and (56), repeated here as (66) and (67):

(66) Direct Complement relatives—Standard or Chopping?
  a. Comi [o bolo(i) [que$_{(i)}$ eu mesma fiz]] *or*
  b. Comi [o bolo(i) [que-COMP eu mesma fiz __$_{(i)}$]] *?*

(67) Subject relatives—Standard or Chopping?
  a. Despachei uma encomenda para [meu tio$_{(i)}$ [que$_{(i)}$ mora na capital]] *or*
  b. Despachei uma encomenda para [meu tio$_{(i)}$ [que-COMP __$_{(i)}$ mora na capital]]

Some authors have argued that subject and direct complement relatives may also present the chopping structure, in a similar way to how indirect relatives do, as shown in the options (b) above; for others, chopping relatives can be analyzed as involving movement in BP (see Kato/Nunes 2014 for a recent review of the debate). As for the copying relatives, empirical data shows that this option is indeed active in subject and direct complement relatives as well (the following two examples are from Kato/Nunes 2014, 581), which might strengthen the analysis in which *que* is not a pronoun in any instance:

(68) Direct object relatives and Subject relatives—Resumptive:
  a. Esse é o livro **que o João sempre cita ele**
  this be.PRES.3SG the book that-REL the João always cite.PRES.3SG it
  'This is the book that João always cites'
  Analysis: Este é [o livro $_{(I)}$ [que$_{(I)}$ o João sempre cita ele$_{(i)}$]] *or*
  [o livro $_{(I)}$ [que-COMP o João sempre cita ele$_{(i)}$]] *?*

  b. Eu tenho uma amiga que ela é muito engraçada
  I have.PRES.3SG a friend that-REL she be.PRES.3SG very funny
  'I have a friend who is very funny'
  Analysis: Eu tenho [uma amiga $_{(I)}$ [que$_{(I)}$ ela$_{(i)}$ é muito engraçada]] *or*
  [uma amiga $_{(I)}$ [que-COMP ela$_{(i)}$ é muito engraçada]]

Similar debates involve other syntactic structures with *que*, perhaps most noticeably, interrogatives (where in fact the conceptual category "movement" translates specifically in a dislocation of the sentence). Note the different positions of interrogative words *como* 'how', in (a) and (b) below (from Hornstein/Nunes/Grohmann 2005, 41–42):

(69) Interrogatives in BP
a. **Como** você consertou o carro?
how you fix.PST.3SG the car
'How did you fix the car?
b. Você consertou o carro **como**?
you fix.PST.3SG the car how

As these examples show, interrogatives in BP may be constructed with the movement of an interrogative word to the beginning of the sentence (as in most Romance languages), but this is not obligatory. More interestingly still, there is a construction using both an interrogative word and the particle *que* (70):

(70) **Como** **que** você consertou o carro?
how QUE you fix.PST.3SG the car
'How did you fix the car?'

The co-occurrence of *que* with interrogative words such as *como* 'how' (but also, *quem* 'who'; *qual* 'which', etc.) in BP raises the question of the nature of this particle as an interrogative pronoun, and in fact as a pronoun at all, in its general use (for instance, in relatives and cleft-constructions, as we have seen). In a more general sense, analysis of the conditions on movement in BP interrogatives have brought to light some interesting problems for the development of generative syntax models, and this remains a topic of interest for researchers in comparative and theoretical fields within this framework.

5 Final Remarks

As we mentioned at the beginning of the chapter, we have tried here to review significant research on the syntax of BP, covering both functionalist and formalist studies; and in order to do so, we have chosen to focus on those aspects of BP syntax that have attracted the attention of those in both fields. This inevitably involved passing over many other important aspects. Some of these are dealt with in the chapters dedicated to Historical Syntax and Morphology, and the reader may also consider some of the more comprehensive accounts cited in the present study: Castilho (2010), Kato/Negrão (2000), Kato/Ramos (1999), and the collection *Gramática do Português Falado* (in particular Kato/Nascimento 2015; Ilari 2014; 2015). We hope that the brief outline presented here may stimulate further interest in the rich literature on BP syntax, within both the functionalist and the generativist frameworks.

6 References

Baker, Mark Cleland (2002), *Syntax*, in: Mark Aronoff/Janie Rees-Miller (edd.), *The Handbook of Linguistics*, Malden, Mass., Blackwell, 265–294.

Berlinck, Rosane/Duarte, Maria Eugênia Lamoglia/Oliveira, Marilza de (2015), *Predicação*, in: Mary Aizawa Kato/Milton Nascimento (edd.), *Gramática do Português Culto Falado no Brasil*, vol. 2: *A construção da sentença*, São Paulo, Editora Contexto, 81–149.

Braga, Maria Luiza (1987), *Topicalizações, deslocamento à esquerda e discurso*, Research report for the CNPQ, Uberlândia, UFU.

Braga, Maria Luiza (1991), *As sentenças clivadas no português falado no Rio de Janeiro*, Organon 5/18, 109–125, https://seer.ufrgs.br/organon/article/view/39126, last accessed 18.02.2022.

Camacho, Roberto (2002), *Construções de voz*, in: Maria Bernadete Abaurre/Angela Rodrigues (edd.), *Gramática do Português Falado*, vol. 8: *Novos estudos descritivos*, Campinas, Editora da UNICAMP, 227–316.

Camacho, Roberto (2013), *Construções relativas nas variedades do português: uma interpretação discursivo-funcional*, Filologia e linguística portuguesa 15/1, 179–214.

Camacho, Roberto (2017), *Alinhamento e estratégias de relativização*, D.E.L.T.A. 33/1, 243–266.

Camacho, Roberto/Pezatti, Erotilde (1997), *Aspectos funcionais da ordem de constituintes,* D.E.L.T.A. 13/2, 191–214.

Castilho, Ataliba Teixeira de (2002), *Gramática do Português Falado*, vol. 1: *Ordem*, Campinas, Editora da UNICAMP.

Castilho, Ataliba Teixeira de (2010), *Nova gramática do português brasileiro*, São Paulo, Contexto.

Coelho, Adolpho (1880), *Estudos para a história dos contos tradicionais*, Revista de Etnologia e Glotologia 1, 108–144.

Cyrino, Sonia (1996), *O objeto nulo do português brasileiro*, D.E.L.T.A. 12/2, 221–238.

Dik, Simon (1989), *The Theory of Functional Grammar*, Dordrecht, Foris.

Duarte, Maria Eugênia Lamoglia (1993), *Do pronome nulo ao pronome pleno: a trajetória do sujeito no português brasileiro*, in: Ian Roberts/Mary Aizawa Kato (edd.), *Português brasileiro: uma viagem diacrônica*, Campinas, Editora da Unicamp, 107–128.

Duarte, Maria Eugênia Lamoglia (2000), *The loss of the "avoid pronoun" principle in Brazilian Portuguese*, in: Mary Aizawa Kato/Esmeralda Negrão (edd.), *Brazilian Portuguese and the Null Subject Parameter*, Frankfurt, Vervuert, 17–36.

Duarte, Maria Eugênia Lamoglia/Kato, Mary/Barbosa, Pilar (2001), *Sujeitos indeterminados em PE e PB*, in: Maria Elias Soared (ed.), *Anais do II Congresso Internacional da ABRALIN*, vol. 1, Fortaleza, Abralin, 405–409, https://www.abralin.org/site/wp-content/uploads/2020/03/ABRALIN_26.pdf, last accessed 18.02.2022.

Faraco, Carlos Alberto (1996), *O tratamento "você" em português. Uma abordagem histórica*, Fragmenta 13, 51–82.

Galves, Charlotte (1989), *O objeto nulo no português brasileiro: percurso de uma pesquisa*, Cadernos de Estudos Linguísticos 17, 65–90, https://periodicos.sbu.unicamp.br/ojs/index.php/cel/article/view/8636801, last accessed 18.02.2022.

Galves, Charlotte (2001), *Ensaios sobre as gramáticas do português*, Campinas, Editora da Unicamp.

Halliday, Michael (1985), *An introduction to functional grammar*, New York, Arnold.

Hornstein, Norbert/Nunes, Jairo/Grohmann, Kleanthes (2005), *Understanding Minimalism*, Cambridge, Cambridge University Press.

Ilari, Rodolfo (2003), *Gramática do português falado*, vol. 2: *Níveis de análise linguística*, Campinas, Editora da UNICAMP.

Ilari, Rodolfo (2014), *Gramática do português culto falado. Palavras de classe aberta*, São Paulo, Contexto.

Ilari, Rodolfo (2015), *Gramática do português culto falado. Palavras de classe fechada*, São Paulo, Contexto.
Kato, Mary Aizawa (1993a), *The distribution of null and pronominal objects in Brazilian Portuguese*, in: *Linguistic Perspectives on the Romance Languages. Selected Papers from the XXI Linguistic Symposium on Romance Languages*, Amsterdam/Philadelphia, Benjamins, 225–235.
Kato, Mary Aizawa (1993b), *Recontando a estória das relativas*, in: Ian Roberts/Mary Aizawa Kato (edd.), *Português Brasileiro. Uma viagem diacrônica*, Campinas, Editora da Unicamp, 223–261.
Kato, Mary Aizawa/Nascimento, Milton (2015), *Gramática do português culto falado: a construção da sentença*, São Paulo, Contexto.
Kato, Mary Aizawa/Negrão, Esmeralda Vailati (edd.) (2000), *Brazilian Portuguese and the Null Subject Parameter,* Frankfurt/Madrid, Vervuert/Iberoamericana.
Kato, Mary Aizawa/Nunes, Jairo (2014), *Uma análise unificada dos três tipos de relativas restritivas do Português Brasileiro*, Sociodialeto 4, 575–590.
Kato, Mary Aizawa/Ramos, Jânia (1999), *Trinta anos de sintaxe gerativa no Brasil*, D.E.L.T.A. 15, 105–146.
Kato, Mary Aizawa, et al. (1996), *As construções-Q no português brasileiro falado: perguntas, clivadas e relativas*, in: Ingedore Koch (ed.), *Gramática do Português Falado*, vol. 6: *Desenvolvimentos*, Campinas, Editora da UNICAMP, 303–368.
Longhin, Sanderléia (1999), *As construções clivadas: uma abordagem diacrônica*, Master's thesis, Campinas, IEL/UNICAMP.
Lopes, Célia Regina dos Santos (1999), *A inserção de "a gente" no quadro pronominal do português: percurso histórico*, Doctoral thesis, Rio de Janeiro, UFRJ.
Lopes, Célia Regina dos Santos (2003), *A inserção de "a gente" no quadro pronominal do portugués*, Frankfurt/Madrid, Vervuert/Iberoamericana.
Lopes, Célia Regina dos Santos (2004), *"Vossa mercê" > "você" e "Vuestra merced" > "usted": o percurso evolutivo ibérico*, Alfal 14, 173–190.
Lucchesi, Dante (2001), *As duas grandes vertentes da história sociolinguística do Brasil*, D.E.L.T.A. 17/1, 97–130.
Melo, Gladstone Chaves de ([3]1975, [1]1946), *A Língua do Brasil*, Rio de Janeiro, FGV.
Modesto, Marcello (2001), *As construções clivadas no português do Brasil: relações entre interpretação focal, movimento sintático e prosódia*, São Paulo, Humanitas.
Negrão, Esmeralda (1999), *Português brasileiro: uma língua voltada para o discurso*, Habilitation thesis, São Paulo, Universidade de São Paulo.
Negrão, Esmeralda/Viotti, Evani (2008), *Estratégias de impessoalização no português brasileiro*, in: José Luiz Fiorin/Margarida Petter/Maria Taddoni (edd.), *África no Brasil. A formação da língua portuguesa*, São Paulo, Contexto, 179–203.
Negrão, Esmeralda/Viotti, Evani (2015), *Elementos para a investigação da semântica do clítico "se" no português brasileiro*, Cadernos de Estudos Linguísticos 57/1, 41–60.
Neto, Serafim da Silva ([3]1976, [1]1950), *Introdução ao estudo da língua portuguesa no Brasil*, Rio de Janeiro, Presença.
Neves, Maria Helena de Moura (2002), *Estudos das construções com verbo suporte em portugués*, in: Ingedore Koch (ed.), *Gramática do Português Falado*, vol. 6: *Desenvolvimentos*, Campinas, Editora da UNICAMP, 201–230.
Neves, Maria Helena de Moura (ed.) (2003), *Gramática do português falado*, vol. 7: *Novos Estudos*, Campinas, Editora da UNICAMP.
Neves, Maria Helena de Moura (2006), *Texto e gramática*, São Paulo, Contexto.
Paixão de Sousa, Maria Clara (2010), *A morfologia de flexão no português do Brasil. Ensaio sobre um discurso de perda*, Estudos da Língua(gem) 8, 55–82.

Pezatti, Erotilde Goreti (1993), *A ordem de palavras e o caráter nominativo/ergativo do português falado*, Alfa 37, 159–178.
Pezatti, Erotilde Goreti (2014), *A ordem de palavras no português*, São Paulo, Parábola.
Pontes, Eunice (1987), *O Tópico no português do Brasil*, Campinas, Pontes.
Rodrigues, Angélica (2009), *Ir e pegar nas construções do tipo "foi fez": gramática de construções de contexto de gramaticalização*, in: Ataliba Teixeira de Castilho (ed.), *História do Português Paulista. Série de Estudos*, vol. 1.3, Campinas, Setor de Publicações do IEL/UNICAMP, 267–278.
Scher, Ana Paula (2006), N*ominalizações em "-ada", com o verbo leve "dar" em português brasileiro*, Letras de Hoje 143, 29–48.
Tarallo, Fernando (1983), *Relativization strategies in Brazilian Portuguese*, Doctoral thesis, Philadelphia (PA), University of Pennsylvania.
Tarallo, Fernando (1988), *Discussing the alleged creole origin of Brazilian Portuguese*, Cadernos de Estudos Linguísticos 15, 137–161.
Tavares, Maria Alice (2008), *Perífrases [V1 (E) V2] em gêneros escritos: propostas para um ensino de gramática baseado no texto*, Linguagem & Ensino 11/2, 329–347.
Torres-Morais, Maria Aparecida (2001), *Resenha de Mary Aizawa Kato/Esmeralda Vailati Negrão, "Brazilian Portuguese and the Null Subject Parameter", Madrid, Vervuert/Iberoamericana 2000,* D.E.L.T.A. 17/1, 155–168, https://revistas.pucsp.br/index.php/delta/article/view/39820, last accessed 18.02.2022.
Votre, Sebastião/Naro, Anthony (1989), *Mecanismos funcionais do uso da língua*, D.E.L.T.A. 5/2, 169–184.

Maria da Graça Krieger

15 Lexicon

Abstract: This chapter addresses the Brazilian Portuguese lexicon from a synchronic perspective. The lexicon is an essential component of a language, defined in traditional terms as the body of its words. The principal goal here is to describe the constitutive heterogeneity of the contemporary lexicon of Brazilian Portuguese (BP). The chapter also outlines the current state of Lexical Studies in Brazil, an area in which lexicology, lexicography and terminology come together. All three areas deal with the lexicon, but differ methodologically in terms of their specific research objects, namely, general lexicon, the dictionary-based lexicon, and technical-scientific terms. The chapter also includes some remarks about Brazilian studies on phraseology.

Keywords: attested lexicon, lexical heterogeneity, lexicography, terminology, Lexical Studies

1 Introduction

> 'The real existence of the unit word in a language makes it possible to consider its whole set, which we call the *lexicon*.'
> (Lara 2006, 143)[1]

A language system is based on two essential components, its lexicon and its grammar. The lexicon is the sum of the words of a language, while the grammar is the system of rules that govern the use of the lexicon in communicative practice. Based on this traditional conception of the two basic pillars that make a language function, this chapter will present an overview of the current lexical composition of Brazilian Portuguese (BP).

The study of a language's lexicon can follow a variety of paths, and its multifaceted nature opens up many horizons for investigation. It would be possible, for example, to take a morphological approach and to look at the basic principles of word formation and at morphosyntactic relations (see ↗13 Morphology). Another, parallel way of approaching the lexicon would be via lexical semantics (see also ↗16 On the history of semantic studies in Brazil), a tendency that has evolved quite fruitfully, in particular in the context of European structuralism. Such views are related to the di-

1 "La existencia real de la unidad *palabra* en una lengua es lo que posibilita y da lugar a la consideración de su conjunto, que llamamos *léxico*."

https://doi.org/10.1515/9783110405958-016

verse aspects of the word and its duplicity of form and content, of signifier and signified.

However, our current orientation privileges a series of aspects related to the heterogeneous composition of the lexical corpus of BP. Heterogeneity is a natural characteristic of the lexicons of natural languages, since they are neither monolithic nor stable. On the contrary, they are diverse in their origins and dynamics, always prone to change and to the acceptance of new words. According to the well-known Brazilian lexicographer Biderman (2001, 178),

> 'The lexicon of any language constitutes a vast universe with imprecise and undefined limits. It includes the whole conceptual universe of that language. Any lexical system is the sum of all the accumulated experience of a society and of the memory of its culture through the generations. The members of this society are the agent-subjects of the process of continuous perpetuation and recreation of their language's lexicon. In this evolutionary process, the lexicon is expanded, modified and sometimes reduced. The social and cultural changes encompass modifications of lexical usage. This is how lexical units or whole sections of the vocabulary can become marginalized, infrequently used, or even disappear. At the same time, however, words can be revitalized and may enter into circulation, generally with new connotations. Finally, new words or new meanings of already existing words emerge in order to enrich the lexicon'.[2]

The changes in the lexicon can also be explained because words have cognitive functions, essential to human communication, since they name, denote, create meaning, and establish reference. They also express concepts, subjectivities and ideologies. Therefore:

> 'the lexicon works as the lungs of the living languages of culture, showing that it is an open set that is renewed, especially, due to its role of naming what is new, what is discovered by science, and the new artifacts resulting from technology' (Krieger 2014, 325).[3]

2 "O Léxico de qualquer língua constitui um vasto universo de limites imprecisos e indefinidos. Abrange todo o universo conceptual dessa língua. Qualquer sistema léxico é a somatória de toda experiência acumulada de uma sociedade e do acervo de sua cultura através das idades. Os membros dessa sociedade funcionam como sujeitos-agentes, no processo de perpetuação e re-elaboração contínua do Léxico de sua língua. Nesse processo em desenvolvimento, o Léxico se expande, se altera e, às vezes, se contrai. As mudanças sociais e culturais acarretam alterações nos usos vocabulares; daí resulta que unidades ou setores completos do Léxico podem ser marginalizados, entrar em desuso e vir a desaparecer. Inversamente, porém, podem ser ressuscitados termos que voltam à circulação, geralmente com novas conotações. Enfim, novos vocábulos, ou novas significações de vocábulos já existentes, surgem para enriquecer o Léxico."

3 "O léxico funciona como o pulmão das línguas vivas de cultura, evidenciando que é um conjunto aberto que se renova, em especial, pelo seu papel de nomear o que surge de novo, o que a ciência descobre, os artefatos que a tecnologia produz."

The functionality of the lexicon explains its naturally heterogeneous formation. Some words decrease in use while others are incorporated as neologisms, regionalisms, or as technical or scientific terms. Such diversity shows that the lexicon is not a static, monolithic block but a changing totality of items and uses built up over time.

The path we take here begins with the relationship between the formation of the lexicon and the general history of the language of which it is a part. The next step will be to offer information on the contributions of lexical studies to our current understanding of the BP lexicon. Finally, the chapter looks at lexicography as a documental source of the Brazilian Portuguese lexicon (BPL).

2 The Brazilian Portuguese lexicon and its constitutional heterogeneity

In order to talk about the BPL, it must be noted that the situation of any transplanted language usually implies a number of problems and issues related both to identity and to the differentiators of the linguistic modality that is spoken in the new geographical space. This is the case with BP in relation to European Portuguese (EP) (see ↗2 The social history of Brazilian Portuguese; ↗8 The debate on Brazilian and European Portuguese; ↗19 Linguistic policy and the orthographic agreement).

As in the case of other languages which exist due to colonial expansion, there is a tradition of treating BP in terms of its contrasts with EP and to mark only the differences. However, the more recent tendency is to consider BP as an autonomous system and to describe it as an independent language. If we look at the differences between Portugal and Brazil, it is only natural that the greatest divergences will be observed in the lexicon, given the crucial role of the word stock in naming things and the resulting dynamics of receiving new words, new expressions and meanings. Hence, one way of describing the gradual formation of the BP lexicon is in fact to begin from a historical and sociocultural perspective (see ↗2 The social history of Brazilian Portuguese).

Brazil was “discovered” in 1500 and was a colony of Portugal until 1822, thus inheriting the colonizer’s language. But this does not imply that Portuguese has always been the only language spoken in the country. During the colonial period, native languages coexisted with Portuguese, Tupi being the most prevalent, which was referred to as *lingua geral* ‘general language’. Many words from native languages merged into BP.

The Brazilian colonial period was also marked by the slave trade, which began in 1502 and resulted in the arrival of African languages, which have been studied systematically only over the past decades: ‘Enslaved people came from different areas

and brought along approximately 200 to 300 languages, mostly originating from the African West and Bantu region' (Petter 2006, 124).[4]

African languages have made their mark on the Brazilian lexicon, especially in the state of Bahia, where most of the Afro-Brazilian population was concentrated, due to the work of the enslaved people in Brazil's agricultural sector there. Religious rituals and food have also been important sources of lexical additions to BP. For example, *Orixás* and *candomblé* are denominations of gods and rituals from African religions that are now familiar words in BP.

The picture of the formation of the BPL also includes the so-called immigration languages (see ↗11 Brazilian Portuguese: contemporary language contacts). The 2010 census conducted by the Brazilian Institute of Geography and Statistics (IBGE) indicates that about 30 languages were brought by immigrants over the last 150 years. Some of these are concentrated in certain regions, such as Italian and German in Southern Brazil or Japanese in the state of São Paulo. Spanish, Polish, Arabic and other languages were brought into other regions. According to the census, some 270 languages result from the sum of native and immigrant languages.

> 'Although most Brazilians have the impression of living in a monolingual country, Brazil is actually multilingual. Within the country, about 200 languages are learned as mother tongues. Brazil's linguistic singularity lies in the fact that one of these languages, namely Portuguese, is currently the language of the vast majority, and all of the others are in the extreme minority. People with minority mother languages in Brazil account for 0.5 % of the country's overall population' (Rodrigues 2013).[5]

The BPL has incorporated many linguistic borrowings. In the first decades of the 20th century, French had an impact here, as reflected in the incorporation of a series of words from French due to the international importance of French culture. Many French words are still widely used today: *menu*, *omelet* in gastronomy, and *boutique*, *lingerie*, *maquillage* 'make up' in fashion.

Currently, English has a considerable impact on the BPL, particularly due to its status as a lingua franca in the international context. Many foreign words have arrived from English, predominantly in technological-based areas and products. Computer technology is the greatest source of foreignisms, and includes examples such as

---

4 "os escravos vieram de áreas distintas e trouxeram cerca de 200 a 300 dessas línguas, basicamente originárias do oeste da África e da região banta."

5 "Embora a maioria dos brasileiros tenha a impressão de viver num país monolíngue, o Brasil é na verdade multilíngue: nele são aprendidas como línguas maternas cerca de 200 línguas. A singularidade linguística do Brasil está em que uma dessas línguas, o português, é hoje extremamente majoritária e as demais são todas extremamente minoritárias. As pessoas que têm línguas maternas minoritárias no Brasil constituem apenas 0.5 % da população total do país". http://www.letras.ufmg.br/lali/PDF/L%C3%ADnguas_indigenas_brasiliras_RODRIGUES,Aryon_Dall'Igna.pdf, last accessed 07.03.2022.

mouse, *tablet*, and *smartphone*. Indeed, the adoption of foreignisms, without any translation, is related to the importation of objects or technological processes that do not exist or are unknown by the community that imports them. In Brazil, these are seldom translated. Media technology has also influenced the use of foreignisms, for example with *podcast* and *streaming*. 'These borrowings come into the language, are integrated in some way into the vocabulary of a linguistic layer, and are modified by the speakers, adapting them to the phonology, morphology and semantics of the Portuguese language' (Timbane/Coelho 2018, 8).[6] Timbane/Coelho (2018) also point out that social media has made a significant contribution to the dissemination of loanwords in the way described above.

The BPL also includes the vocabulary of borderlands, which suffer the influence mostly of the Spanish language, since Brazil is a Lusophone island surrounded by Spanish-speaking countries (see ↗11 Brazilian Portuguese: contemporary language contacts). This scenario shows the constitutional heterogeneity of the BPL in terms of its linguistic origin. However, its base still is the considerable percentage of words from Latin origin due to heritage of Portuguese as a Romance language.

For all these reasons, BP constitutes a real lexical kaleidoscope; the mix includes general lexical items, shared by all, as well as items whose use is regional, and thus typical of internally delimited geographical spaces. An example of this is the result of a loan from German dialects spoken in Southern Brazil. The noun phrase *ein krenke* 'a sick person', led to a word that refers to a difficult situation: *encrenca*. This word is used frequently in this region.

However, since almost 100 % of the population, including immigrants, speaks Portuguese, the communication between Brazilians is not a problem and presents no major conflicts, and Brazil has adopted Portuguese as its official language, a situation which persists to this day (↗19 Linguistic policy and the orthographic agreement).

The latest agreement on orthography (1990), which sought to unify spelling across the nations that use Portuguese as an official language, now relies on an important parallel project that seeks to pave the way for a wider cross-referencing of information on the shared lexicon. The lexical records that make up the *Vocabulário Ortográfico Comum da Língua Portuguesa* 'Common Orthographic Vocabulary of the Portuguese Language', known as VOC (Instituto Internacional da Língua Portuguesa s.a.), constitute, according to the Brazilian linguists involved, not only the basis for the application of the agreement across the 'Community of Portuguese Speaking Countries' (*Comunidade de Países de Língua Portuguesa*, CPLP), but also an 'unprecedented advance in the lexicography of the Portuguese language'[7] (Almeida et al. 2013, 205).

---

6 "Esses empréstimos 'entram' na língua, integrando-se de alguma forma ao vocabulário de algum estrato linguístico, sendo que o falante modifica, aportuguesando-as fonológico, morfológico e semanticamente na língua portuguesa."

7 "Um avanço sem precedentes na lexicografia de língua portuguesa."

The VOC has recorded more than 300,000 entries and over 2 million orthographic forms. It is also a gateway to a more detailed knowledge of all the spoken and written variants of the language in the CPLP member-countries. The cross-referencing of data will also contribute to a better recognition of the lexical configuration of the Brazilian variant of Portuguese. In this way, the VOC mirrors the current research on BP and its lexical component.

Also to be considered here is the 'Orthographic Vocabulary of the Portuguese Language' (*Vocabulário Ortográfico da Língua Portuguesa*, VOLP), developed by the Brazilian Academy of Letters (Academia Brasileira de Letras [6]2021). VOLP's latest online version (2021) has 381,000 lexical records. Each record corresponds to a *type*, a criterion that identifies each word in a language according to the terminology and principles of word occurrence in Corpus Linguistics.

Standard language and literary expression are the original bases sed in VOLP. Recently, however, the Brazilian Academy had started to include words and expressions that are widely used in the media and by the population in their daily lives. Some recent examples are *feminicídio* 'femicide' and *antirracista* 'anti-racist'.

Another source that contributes significantly to the knowledge of essential aspects of the BPL is a line of research that we might refer to under the umbrella term Lexical Studies, and which includes lexicology, lexicography, and terminology.[8] With their specific fields of work, these three sub-areas aim to identify the attested lexicon in its various configurations, as we will discuss in the following section.

3 Lexical Studies: lexicology, lexicography, terminology

In the three mentioned disciplines—lexicology, lexicography and terminology—, the concept of lexicon is understood in terms of its traditional definition, that is, as the total stock of words in a language.

These disciplines deal with the attested lexicon, but each defines its identity by means of its particular objectives and research aims. Lexicology is accepted as the scientific study of the general lexicon; lexicography is the art and technique of producing dictionaries in one or more languages; terminology prioritizes technical-scientific terms. Lexicology and terminology focus on the semantically full grammatical classes, such as nouns, adjectives and verbs, which are the basis of the vocabularies

8 Over recent decades, this trio of research areas has attained *bona fide* status in the Brazilian Association of Post-Graduate Studies in Letters and Linguistics. This Association gathers graduate students and researchers and studies into thematic groups, among which the Grupo de Trabalho de Lexicologia, Lexicografia e Terminologia 'Working Group on Lexicology, Lexicography and Terminology'—GTLEX is among the oldest. http://www.letras.ufmg.br/gtlex/, last accessed 18.02.2022.

studied. On the other hand, lexicography, which aims to register the lexical component as a whole, covers all grammatical classes.

This set of three disciplines can be classified as Lexical Sciences, a specific area of study in Brazilian research. The principal themes of lexicology are neologisms, toponyms, regionalisms, among others. Research on neologisms (see also ↗13 Morphology) shows the dynamism of the lexical component, as well as how it welcomes new words and expressions. New lexical items are coined in order to name new phenomena and to project new meanings into linguistic communities. The identification of neologisms varies according to the texts and genres under examination (see ↗17 Spoken *vs*. written language). The most prevalent are a variety of texts from journalism, advertising, and comedy, including cartoons that combine verbal and non-verbal elements. Notable studies on the formation and behavior of lexical neologisms within BP are based on data from the Neologism Observatory at the University of São Paulo.[9]

Data collected over a period of twenty years shows, for example, that 'prefix derivation has remained the most productive process in forming neological formations' (Alves 2010, 67).[10] Among numerous other aspects, subtle alterations have also been detected in the use of prefixes that have a tendency to express intensity, such as *super*, *hiper* and *maxi*. The prefix *super*, for example, is especially prevalent: it attaches to different word classes, forming nouns such as *supermercado* 'supermarket', *superaliança* 'super-alliance', superlative forms of adjectives such as *super elegante* 'very elegant' and verbs such as *superdimensionar* 'oversize' (Alves 2010).

The sum of all these studies into neologisms well illustrates the ways in which BP renews itself over time. No alterations are sufficiently profound, however, to question the status of BP as a variant of the Portuguese language, since neologisms, whether foreign or vernacular, adapt themselves to the given morphosyntactic standards.

Another productive strand of lexicological research is that of *toponyms* (see ↗18 Onomastics and toponomastics), although proper nouns as identified in the grammatical tradition are not part of the general lexicon of linguistic communities. Nevertheless, within onomastic research, the processes involved in naming places have become the subject of a toponymic theory that categorizes them according to a typology, thus identifying modes and types of naming (Dick 2001). This is achieved by considering aspects that relate to both local history and ethnicity, such as certain types of reiterated formants, as e.g., *polis* in the town names *Teresópolis* and *Petrópolis*.

The same categorization has been used in hydronymic studies that seek to show the relationships between place names and their natural environment. *Pantanal* 'swampland', a region of Central Brazil and a route for those in search of gold and

9 http://ccint.fflch.usp.br/observatorio-de-neologismos-do-portugues-brasileiro-contemporaneo, last accessed 18.02.2022.

10 "a formação de neologismos por meio da derivação prefixal tem-se mantido como o processo mais produtivo na formação de novas unidades lexicais neológicas."

other precious metals, '[...] is an example of the toponym of one of the largest floodable areas on the planet'[11] (Isquerdo/Seabra 2010, 79).

Based on African words such as *banana* and *macaco*, examples of this influence include geographic names such as *Córrego Bananal* and *Cachoeira do Macaco*, among numerous others (Isquerdo/Seabra 2010, 92). It is widely accepted that toponyms have been an important descriptive source arising from the influence of economic interests, such as gold prospecting and mining, as well as factors of ethnicity, in the formation of Brazil's overall lexicon.

The regional lexicon is dependent on historical, political and cultural factors. This explains why the word *chimia*, a kind of jam, is used in only one region. It was introduced by German migrants in Southern Brazil and is unknown in other parts of the country. However, for some terms, there are variants that are nationally known, such as the classical example of the manioc root used in Brazilian cooking, with three denominations. In the Southern and Southeastern region of Brazil, it is called *aipim*; in the Northeast, *macaxeira*; in the Midwest and part of the North, *mandioca*.

A systematic identification of Brazilian speech in terms of vocabulary is the aim of the Brazilian Linguistic Atlas (Cardoso/Mota 2012; ↗9 Dialectology and linguistic geography). Despite its lexicological character, the project is considered to belong to the field of dialectology and is developed based on regional atlases, some of which have already been completed. The complete publication of the Atlas will enable researchers to identify distinctive regional speech patterns, including several important data for a better understanding of the BP lexicon.

Likewise, monolingual lexicography is an essential source of lexical data. Standard dictionaries offer a great deal of information about the words and expressions they record. In addition to meanings, they provide etymological and sociolinguistic information related mainly to regional uses. Brazilian lexicography is thus an important point of reference for studies of BPL. In particular, the main Brazilian dictionaries published in the 20th century allow us to verify that the identification of the Brazilian lexicon occurs as a mark of national identity as opposed to European Portuguese (EP). It is a late and specific acknowledgement of aspects of the nature and cultural reality of Brazil, as will be discussed later in this chapter.

Another contribution of lexicography is that it emphasizes the critical examination of the production of dictionaries. The works under investigation here are not only both older and contemporary monolingual ones, but also bilingual dictionaries, due to the importance of these for the acquisition of foreign languages, in addition to their relationship with translation. With such a variety of focuses, theoretical lexicography branches out in many directions, and never really establishes a uniform analytical paradigm (Krieger 2020).

11 "[...] é um exemplo de topônimo da maior área inundável do planeta."

However, linguistics offers solid foundations for the study of elements to be found in many types of dictionaries. The Lexicon Sciences[12] series brings together analyses and reflections on the constitution of nomenclatures, the quality of definitions, the organization of lexicographic articles, the choice of examples and accreditations, among other topics.

Even so, a convergence of theoretical principles can be identified in the field of what is referred to as Learner's or Pedagogic Lexicography, an area that focuses on foreign language learning from the earliest stages of acquisition onwards, in which the role of dictionaries is essential. In general, looking up words is restricted to confirming their existence, orthography and meanings.

Therefore, some lexicographic studies in this field have intensified their activities, on two fronts: first, to establish criteria for a proper lexical focus in teaching a mother tongue, and second, to make clear that the '[monolingual] dictionary is a privileged place for lessons on language'[13] (Krieger 2012, 9). This latter concept is the case due to the copious information that the dictionary offers concerning the lexicon, its use, and meanings. This defines its didactic nature and turns it into a potential learning tool when teaching/learning a first language.

These principles have had a major impact on the concept of school dictionaries, which are generally mistaken for mini-dictionaries. In 2006, a typology of school dictionaries was proposed to the Brazilian Ministry of Education, based on a perceived relationship of adequacy between the type of work involved and the linguistic level of learners. This classification has actually been adopted by the Ministry of Education and has gone on to serve as a guide for the 'National School Book Program' (PNLD) (Krieger 2008), which selects and purchases school books and dictionaries for use in Basic Education, that is, for the first eight years of public education.

Concurrently, the foundations of Mother Tongue Pedagogical Lexicography have led to a broad field of theoretical research in Brazil and practical guidance for elementary education teachers. In fact, over the last twenty years, studies on different types of dictionaries have intensified, thus contributing to a consolidation of an important theoretical dimension in lexicography and assuring its recognition as a field in applied linguistics.

Likewise, the research field of terminology, the youngest of Lexical Studies, has significantly expanded in Brazil over the last twenty years. Terminological studies have contributed to the identification of specialized lexical components, the recogni-

12 The collection *As Ciências do Léxico: Lexicologia, Lexicografia, Terminologia* has been published since 1998, organized by Aparecida Negri Isquerdo from the Federal University of Mato Grosso do Sul under the auspices of the GTLEX group. It integrates works from ANPOLL researchers of the respective specializations. Nine volumes have been published thus far. In 2020, volume IX was launched in eBook format https://www.ufms.br/lancamentos-da-editora-ufms-abordam-lexico-literatura-comunicacao-e-educacao/, last accessed 14.03.2022.

13 "[...] o dicionário de língua é um lugar privilegiado de lições sobre a língua e a linguagem."

tion of which is always complex for two main reasons, the conceptual dimension of terms as well as their formal composition.

The first aspect can be explained in that a lexical unit is a term when it expresses a concept in a specialized scientific or technical field such as chemistry or computer science. 'For specialists, terminology is a formal reflection of the conceptual organization of a specialized area, and an indispensable means of expression and professional communication' (Cabré 1993, 37).[14]

The lexical unit *casa* 'house' is defined by common sense as the place where someone resides. However, *casa* is also defined as the 'inviolable asylum of the individual'[15] in the Brazilian Federal Constitution (Senado Federal 2016, article 5, item XI). Thus, *casa* is a legal term. Unlike the word itself, a term does not exist *a priori*, but only acquires terminological value in specialized communication. Most words and terms are formally similar lexical items. The main difference between them is of a functional nature, because terms are explicitly defined from the point of view of their designation, and thus are essential items for professional communication.

The second aspect is the formal composition of the terms: there are lexical units that are not exclusive to terminologies, such as *casa*, which is part of the common lexicon. However, there are terms that are coined based on the tradition of technical-scientific nomenclatures, structured with Greco-Latin formants. In medical terminology, this tradition can be seen in the examples *nefropatia* and *cardiologia* 'nephropathy', 'cardiology'. In law, terms such as *jurisprudência* 'jurisprudence', and in biology, *patógeno* 'pathogen, a specific agent that causes disease', are other examples of the same kind of terminological formation that cannot be mistaken for the common lexicon. Terms with this type of formation seek an ideal univocity because they favor the understanding of a single concept, avoiding ambiguity of meanings, which is fundamental in specialized communication.

In addition to this, the difficulty in identifying terminology occurs because the terms are rarely equivalent to one single lexical item. Terminological syntagms are common, as in the examples *desenvolvimento sustentável* 'sustainable development' and *plantas geneticamente modificadas* 'genetically modified plants', terms related to Brazilian environmental laws.

The syntagmatic composition of terms has broadened the view of the field of terminology, which has gone on to describe specialized phraseologies. Lexical compounds of this kind have been the object of intense investigation. In this respect, the proposals of the group 'Southern Cone Terminology Project' (Termisul) at the Federal University of Rio Grande do Sul, coordinated by Cleci Regina Bevilacqua, stand out.

---

14 "Para los especialistas, la terminología es el reflejo formal de la organización conceptual de una especialidad, y un medio inevitable de expresión y de comunicación profesional."

15 "asilo inviolável do indivíduo".

One of the main focuses of its research is specialized lexical combinations, present in legal, normative and scientific language (Bevilacqua/Reuillard 2010, 433).

Technological research has also emphasized the importance of lexicon/text relations as a tool to approach terminologies. This is due to the impossibility of specialized communication in the absence of technical terms; therefore, the text is regarded as the natural habitat of terminology. The focus referred to as textual terminology has favored a dialogue with the linguistics of specialized languages (Finatto/Zilio 2015).

At the same time, terminological identification has moved towards current technological resources, following the methodology of Corpus Linguistics. This methodology has been developed by the GETerm[16] group and speeds up terminological recognition, considering the high frequency of complex lexical units. Results from research of this kind contribute to the expansion of the applied facet of the study of Terminology, in the production of glossaries, technical dictionaries and terminological databases. Indeed, some of these are developed by terminology researchers themselves. Whether in theoretical or applied terms, research here encompass a variety of scientific, technical and technological areas, in addition to others that are not traditionally investigated, as is the case with other terminological categories identified, such as public and private administration.

Several researchers in the area of terminology have also developed dictionaries and glossaries in specialized fields.[17] In general, works dealing with terminological references tend to indicate equivalent terms in other languages, hence terminology and translation theoretical studies are in a constant dialogue. The results generated in the area of lexical studies have contributed to better understanding of the composition of the attested BPL from the perspective of neologisms, regionalisms, toponyms, technical and scientific terminologies, among other components. Based on theoretical and methodological proposals according to specific objectives, lexicology, lexicography and terminology collectively show how dynamic and heterogeneous the formation of the lexical component can be. Thus, this trio reiterates its identity as having been shaped within the limits set by the study of the lexicon as a body of attested words in a language. As we have said, lexicology and terminology have the essential aim of describing semantically full lexical classes: nouns, adjectives, and verbs. On the other hand, lexicography, which covers all grammatical classes, allows the identification of the general profile of the BPL as a whole, as we will discuss in the following section.

16 Grupo de Estudos de Terminologia da Universidade de São Carlos. http://www.realiter.net/pt/presentazione/membri/geterm, last accessed 14.03.2022.

17 *Dicionário de Linguística da Enunciação* (Flores et al. 2009); *Dicionário de direito ambiental: terminologia das leis do meio ambiente* (Krieger et al. 1998); *Glossário de termos neológicos da economia* (Alves 2001).

4 Lexicography

The broad identification of a language's lexicon is usually associated with the lexicography of a particular country. Thus, it is understood that the traditional dictionary of the language—the most prototypical among lexicographical works—essentially gathers the body of lexical items created and used by one linguistic community over the course of its history. This type of dictionary is, therefore, the greatest testimonial to the constitution of a language's lexicon.

The process of making a dictionary from the lexicon of a language, with the identification and definition of the items that comprise it, with the register of its various uses, examples, and the main grammatical characteristics of each lexical item, mean that the dictionary is a "text" with its own set of rules, and not merely a list (cf. Haensch/Wolf 1982).

The list of lexical entries constitutes the macrostructure of a dictionary, while the entry itself, with its many pieces of information, corresponds to the microstructure. This organizational and informative set makes the dictionary 'a specific, textual, metalinguistic, cultural object' (Rey 1977, 6).[18] Rey adds that the dictionary provides us with a picture of the lexicon. Such a concept is supported by the fact that it is difficult to make a complete register (rather than a partial picture) of all the words and expressions used by a linguistic community.

There are always differences between dictionaries; they are not all created equal, despite the socially disseminated belief that such works provide full objectivity in identifying and defining the lexical items of a language. Lexicography involves identifying items, an area that might involve different propositions and methods, even though the purposes are often shared. The common and basic methodological principle of which words to include depends on their frequency of use. The definition of the meanings remains in the background. Then, as we move from word to meaning, lexicography is characterized as semasiological.

On the other hand, the history of a country's lexicography reflects the path of formation of its linguistic identity. Since its origins, Brazilian lexicography (see also ↗7 The history of the lexicon) has always been accompanied by the issue of BP identity in opposition to EP. Early on, this was confused with Portugal's identity, since the first dictionaries recording the BP lexicon were in fact from Portugal. Although the Portuguese arrived in Brazil in 1500, a national Brazilian lexicography was only established and consolidated in the 20th century. This late appearance is associated with the fact that publishing was forbidden in Brazil, due to its status as a Portuguese colony (see ↗2 The social history of Brazilian Portuguese). The freedom to publish came with the country's independence in 1822.

18 "un objet spécifique, textuel, métalinguistique, culturel".

> By the 20th century, which was a landmark for Brazilian lexicography, there was a considerable gap to be filled, although there were a few pioneering initiatives. Previously, there had only been small leaflets intended to serve as complements to Portuguese dictionaries. The systematic recognition of the lexical identity of BP is expressed especially in a body of six works that are regarded as the founding dictionaries of Brazilian lexicography (Krieger et al. 2006).

The first records of the BPL appeared in two prestigious Portuguese dictionaries that started publishing joint Portugal/Brazil editions: *Novo Dicionário da Língua Portuguesa* (Figueiredo [14]1949) from the time of its fourth edition in Portugal, and *Dicionário Contemporâneo da Língua Portuguesa Caldas Aulete*, which was published in Brazil in 1958 and to this day continues to publish updated editions. Both publications already had impressive word lists for the time, with approximately 130,000 entries.

The joint effort of publishers from the two countries is largely explained by the fact that publishing in Brazil implied a substantial expansion in the market. Although very small, the inclusion of Brazilian terms expresses the recognition that no thesaurus of the Portuguese language could be complete without including lexical items from BP, which were then regarded as "popular vocabulary".

The issue of EP versus BP is a constant in the essentially Brazilian lexicographic cycle, which began with *Pequeno Dicionário Brasileiro da Língua Portuguesa* (Barroso et al. 1938), the first dictionary whose title states its nationality. This dictionary was created with the clear objective of recording the BP lexicon, and it effectively did so in a succession of editions, until its disappearance in the 1980s. It became a renowned reference book and came to include as many as 60,000 entries in a single volume. For the first time ever, items from the mix of indigenous and African languages used in the country were recorded.[19]

As a large-scale work, the *Grande e Novíssimo Dicionário da Língua Portuguesa* (Freire 1939–1944) was published gradually. This was considered to be the first large Portuguese language dictionary printed in Brazil, published in five volumes. It was the first dictionary to aggregate words that had already been listed in dictionaries from Portugal with brazilianisms and Brazilian regionalisms. Despite this, the dictionary received harsh criticism because it failed to formalize the presence of brazilianisms. In the words of the author himself:

> 'Created mainly for Brazilians, this dictionary requires no indication of brazilianisms for the knowledge of the language spoken within the country. In addition, it is not an easy task to define what a brazilianism is' (Freire 1939–1944, VIII).[20]

---

19 *Pequeno Dicionário Brasileiro da Língua Portuguesa* (Barroso et al. 1938) is a collective work that was organized, for the first few editions, by José Baptista da Luz, Hildebrando de Lima, Gustavo Barroso and Aurélio Buarque de Holanda Ferreira.

20 "Feito principalmente para Brasileiros, este dicionário não precisa de indicação de brasileirismo para conhecimento da linguagem falada no país. Além disso, não é fácil definir o que seja brasileirismos."

Indeed, the topic of brazilianisms can be found across the board in Brazilian lexicography, whose development was one of the objectives at the inauguration of the Brazilian Academy of Letters in 1897. This goal was only achieved with the *Dicionário da Língua Portuguesa* (Nascentes 1961–1967), a publication that did not achieve the influence that the Academy hoped for and did not manage to cover the written and spoken body of the Brazilian lexicon.

The work that finally gained wide visibility and became the benchmark for the BPL was the *Novo Dicionário Aurélio da Língua Portuguesa*, by Aurélio Buarque de Holanda Ferrreira. The first edition dates back to 1975 and was already characterized at that time as a large publication, due to the size of the word list, which came close to 150,000 entries. In following editions, this would increase to approximately 435,000 words.

The greatest merit of the *Aurélio*, as Brazilians tend to call it, is the thorough record of almost 50 % brazilianisms within its entries. It identifies lexical units that originated in Brazilian culture and which were circulating within the country. In addition, it marks regionalisms, indicating the Brazilian region or state where the unit was created and/or used in a particular manner, since differences here might be only of semantic nature. This pioneering work has become a standard reference for the uses and meanings of the language spoken and written in the country. This is why the *Dicionário Aurélio* has ultimately assumed a prominent position among the founding dictionaries of Brazilian lexicography, standing practically alone in the world of lexicographic publishing during the second half of the 20th century. In summary, the *Aurélio* dictionary unequivocally establishes the country's lexicography, whose identity at the beginning of the process was marked by the records of BP.

In fact, this group of six founding dictionaries constitutes an essential source of information regarding the conception of the BPL. Such a conception can be identified based on the records and data included in each lexical entry, since the front matter of dictionaries does not specify the theoretical and methodological principles used to classify brazilianisms.

However, an analysis of the marks indicating brazilianisms and regionalisms in these dictionaries shows that, even though the six works are in themselves divergent records to some degree, an interesting regularity can be observed in lexical items formally indicated as being Brazilian: flora, fauna, food, and native tribes. Some examples of this thematic prevalence are the fruits *abacate* 'avocado' and *araçá*; the animals *araboia* and *tabarana*, denoting a snake and a fish, respectively; likewise, food items, such as *arabu* and *tacacá*, and the name of indigenous tribes, such as *Tabajaras* are prevalent (Krieger et al. 2006).

The dominant themes are representative of words related to the physical nature of the country and its most typical cultures, which did not exist in Europe. Within these themes, there is a prevalence of lexical items from Tupi, the most widespread indigenous language in Brazil, but there are also records of words from African languages. In the field of food, there are also new names borrowed from the various ethnicities that found themselves coexisting in Brazil.

Therefore, differences in lexicalization pertaining to physical, social, and cultural realities appear to be the overall criteria to outline the boundaries between the lexicon of EP and the Brazilian variant, according to the records of the founding dictionaries of Brazil's lexicography.

Currently, dictionaries are no longer shaped by a concern for recording a typically national lexicon, although they continue to do so. Their horizons have broadened, and are no longer limited to regarding the lexicon solely through the lens of the Portugal/Brazil opposition (↗8 The debate on Brazilian and European Portuguese). This is the case with the *Dicionário Houaiss da Língua Portuguesa* (Houaiss 2001). A posthumous work named in honor of the renowned Brazilian philologist Antonio Houaiss, it is characterized by records that cover a greater diversity of modalities of the Portuguese language, in an attempt to account for the language as a whole. The geographical reach of the Portuguese language is reflected in the extent of this work by Houaiss and his collaborators, which attempted to account for the Portuguese of the CPLP as a whole. The first edition includes 228,500 lexical units, without privileging specific chronological stages.

In addition to including painstaking grammatical information on each entry, the *Dicionário Houaiss* is unique in its lexicographic record of the history of the formation of the Portuguese lexicon since it includes dates and etymology for each of the attested lexical items. Additionally, the dictionary often identifies the origin of the word from a regional standpoint. Despite its great value in terms of knowledge of the language, the *Dicionário Houaiss*, like the others discussed here, fails to specify the reference sources that have informed the compilation of its entries.

Unlike those already mentioned, one aspect of Brazilian lexicography has been innovative in various ways, including the overt statement of its organizational principles and criteria. This is the case with two publications: the *Dicionário de Usos do Português do Brasil*, the first edition of which was published in 2002 (Borba 2002), and the *Dicionário UNESP do português contemporâneo*, published in 2004 (Borba 2004). Both works are by Francisco da Silva Borba and his highly qualified team of collaborators. The former is presented as a dictionary of the written language in Brazil in the second half of the 20th century. Its objectives include providing elements to assess the syntactic-semantic properties of the lexicon and fostering reflection on the use of the language. The latter has similar objectives and is targeted basically at secondary school students.

There are two main innovative characteristics in these two works: the first is that both define and apply theoretical-methodological principles and foundations of monolingual lexicography; second, they use an electronic corpus to compile their respective nomenclatures. The former aspect justifies the emphasis on the theoretical dimension of lexicography, which, in this sense

> 'seeks to establish a body of principles that enable the (total or partial) lexicon of a language to be described, thus developing a metalanguage to manipulate and present the relevant information. It is in this aspect that the contribution of linguistic theory comes into play' (Borba 2003, 15).[21]

In addition to the linguistic principles that govern the treatment of the data recorded, the second innovative aspect here is the methodology of selecting the body of entries of the dictionary based on a corpus. The use of a corpus also serves as a foundation to build the word list of the *Dicionário de Usos do Português do Brasil* (Borba 2002). The lexical base for this work includes 70 million items, with words that can't be found in more traditional dictionaries. It includes references for 62,000 entries and reflects the most commonly used words in contemporary BP from the 1950s onwards. With the same corpus methodology, the *Dicionário UNESP do português contemporâneo* includes information about some 90 million lexical items in texts written in Brazil from 1950 onwards (Borba 2004).

The development of these dictionaries has been supported by a database designed and structured by linguists and IT professionals at the Lexicography Laboratory of the Sciences and Letters College in São Paulo State University (UNESP), in the city of Araraquara. A pioneering lexicography center in Brazil has also been established, one which can enable the development of other dictionaries that have followed the corpus methodology in determining their lexical bases: *The Brazilian Corpus*, regarded as the largest of its kind, is also hosted by a university (PUCSP), and is coordinated by Antonio Berber Sardinha (s.a.). It consists of a collection of approximately one billion words from contemporary BP. The methodology for the use of the corpus also made possible the composition of the text archive that served as the basis for the *Dicionário Histórico do Português do Brasil* (DHPB) which accounts for the 16th, 17th and 18th centuries.[22]

This dictionary registers 10,470 entries, all attested through documents of various genres, dating back to three centuries of Brazilian colonial history. It is a documental work, and all the definitions include their context, together with the earliest date in which the entry word was used. Its data thus offers a clearer picture of the Brazilian lexical repertoire in the country's first centuries of development.

As already noted, the most recent examples of contemporary Brazilian lexicography have changed paths and methodologies. No longer are works developed with the aim of identify the lexicon that is typical of BP, which has already been historically es-

---

21 "procura estabelecer um conjunto de princípios que permitam descrever o léxico (total ou parcial) de uma língua, desenvolvendo uma meta linguagem para manipular e apresentar as informações pertinentes. É nesse aspecto que entra a contribuição da teoria linguística."

22 The *Dicionário Histórico* ('Historical Dictionary') was designed by Maria Tereza Biderman and concluded after her death by Clotilde Azevedo Murakawa (Biderman/Murakawa 2021). The dictionary was developed in the Lexicography Laboratory of the Letters and Science College, São Paulo State University, Araraquara.

tablished. This is largely due to the dictionaries that have become classics at a national level, either because of the extent of their entry lists or due to the consistency of their treatment of data. They are all known by the names of the lexicographers who have conceived them: *Aurélio*, *Houaiss*, and *Caldas Aulete* remain household names in the domain of language dictionaries, but have also adopted new platforms to disseminate and facilitate access to their entries. Most dictionaries, from classics to school dictionaries, are available in electronic versions or via online versions.

The greatest impact of digitalization on lexicography in Brazil and globally has been the expansion in the number of lexical items recorded. The criteria adopted diverge from the historic tradition of lexicography. For instance, the number of entries of *Caldas Aulete Digital* (the digital version of the dictionary) has doubled since its last physical version (2004).

The goal of assuring that the words included constitute an accurate reflection of the current linguistic reality influences the selection of documental sources. The Internet is key here, offering a dynamic universe of lexical items stemming from communication through social networks, which become increasingly frequent, to the point that they necessarily become incorporated into dictionaries. This includes a mix of lexical items from various origins and languages.

In this context, the number of foreign words used in the everyday linguistic practices of Brazilians has grown. Examples of this include *selfie*, *hashtag*, and many others. The use of foreign expressions can be quickly measured and incorporated into electronic dictionaries, which are more dynamic than their paper-based counterparts. Also, globalization has contributed to blur what were once rigid linguistic borders, a factor that facilitates international communication. Despite this, the impact of foreign words on current BP is far from having reached damaging proportions.

Other impacts on the lexical repertoire arise from the great scientific and technological advances that also influence people's everyday lives. Given its nature in naming things, the lexicon expresses all of this scientific, technical and technological development, causing lexical items to express not only general meanings but also concepts from scientific and technical areas. The status of being a technical-scientific term is generally indicated by the network of meanings of the lemma, marked with the respective area of specialization. These kinds of records, called usage marks, have developed in Brazilian lexicography, as we can see in the case of the word *rede* 'net', which in the field of electricity denotes the 'distribution circuit of electric current' (Ferreira [2]1986, 1466).[23]

Electronic dictionaries, which are also available as mobile phone applications, are characterized as being usage-based, in addition to working as dynamic tools for linguistic reference, moving away from the paradigm of the slow incorporation of neologisms established by traditional lexicography. By contrast, electronic dictionaries

23 "Eletricidade. Circuito distribuidor de corrente elétrica."

recognize real, contemporary language use and naturalize lexical items within a short period of time, especially those that circulate on social networks, such as *printar* 'print screen' or 'print' and *tuitar* 'tweet' on Twitter.

This brief overview of Brazilian lexicography, which began in Portugal and gradually became geographically national to Brazil itself, has described the paths and stages that led it to its current state of autonomy. Its beginnings were marked by the aim of recording the BP lexicon. This was a difficult task, given the common use of the Portuguese language, which had been transplanted by the first Europeans to arrive in Brazil. Another early challenge was identifying points of conjunction and disjunction between European Portuguese and its Brazilian variant. However, this concern slowly disappeared over time.

Despite this, lexicography in Brazil has laid the foundations for differentiating the Brazilian lexicon from that used in Portugal. With even those few differences in register to be found in the founding dictionaries, common themes were developed to account for a general profile of the BPL. Items related to the flora, fauna, names of indigenous tribes, food items and folklore were thus marked as brazilianisms. Etymological information, including indigenous and African languages, has also been a determining factor in the lexicographic categorization of brazilianisms.

A new version of the lexical recognition of BP follows the principles of Corpus Linguistics. The notion of occurrence has thus come into play, associated with the frequency of use of words and expressions. In this way BP has come to be recognized in a huge body of diverse texts systematically selected in order to be representative of the lexical universe that exists within the country. Hence, the data compiled in dictionaries is now based on frequency of use and not on the ideal of a correct and proper form of speech, as was the case with classical lexicography. Contemporary lexicography, then, reflects a renewal of BPL uses and confirms the dynamic lexical nature of this field.

5 Multivocabulary units and phraseology

The task of preparing a monolingual dictionary also requires dealing with other difficult problems such as identifying multi-word lexical units. In BP, as in other languages, the lexical units have different composed structures recorded in the nomenclatures of Brazilian dictionaries. For example, *guarda-chuva* 'umbrella' and *couve-flor* 'cauliflower' have their own entries due to the hyphenation in their structures. This is the formal mark of syntagmatic consolidation of complex lexical units in BP.

A non-hyphenated multi-word unit will appear more frequently as a sub-entry if its composition includes the lemma of the entry. Normally, the sub-entry is graphically highlighted and positioned after the network of primary meanings, as in *cabeça dura* 'hard-headed', integrating the word *cabeça* 'head'. The example shows that the choice here includes stereotypical BP expressions. Recording them has become an absolute necessity, and it is also compatible with a more contemporary perspective:

> '[...] a lexicography based fundamentally on communication, which stems from the intrinsic value of the vocabulary in the communication process, of the modes and situations in which a lexical unit is used within a linguistic community' (Haensch/Wolf 1982, 19).[24]

In Brazil, the use of popular phraseological units, such as *pisar na bola* 'to mess up' and *dobrar a língua* 'to watch your mouth', is quite common. In both examples, the meaning transmitted does not depend on decodifying each word. The semantic autonomy of the components of these structures and their frequent use indicate that they have acquired new properties through lexicalization. This justifies their inclusion in dictionaries, even if they do not figure in the set of entries itself. They are, rather, recorded as sub-entries within other entries when the lemma is a word in the registered phrase. Hence, *dobrar a língua* is a sub-entry of *língua* (tongue). This is a practical solution that meets the semasiological principle of lexicography.

In most cases, although they take on semantic and pragmatic autonomy, multivocabulary structures and phraseological units are incorporated into dictionaries as additions to the microstructure of simple lexical items. The lexicon thus expands structurally, but is anchored in simple words, according to the forms of lexicographic register.

For a long time, simple lexicon items were viewed as "pure" units, which represented the lexicon *par excellence*. Multivocabulary units disturb this concept of lexicon. They take on the value of a word, expressing a unified, whole meaning, which causes its lexical components to become semantically opaque. Thus, the "impure" units work as lexical elements in communication, exactly like simple units.

However, the lexicon itself is composed of simple and syntagmatic items, never of phrases. The presence of phraseological units is compounded by factors of meaning. It is important to note that monolingual works, alongside the image of the lexical component of languages, offer information on their common uses in the speech of linguistic communities. Due to this, the metalinguistic nature of dictionaries is confirmed, as well as the reason for their being referential works of the lexicon and its uses.

Multivocabulary structures are a complex issue in several respects, starting with the denominational and conceptual differences involved. Locutions, collocations, fixed expressions or idioms, phraseological or conventional units or stereotypes all form part of a multi-faceted universe in which there is no conceptual unanimity in terms of theoretical schools of thought.

The study of these linguistic structures has received a great deal of attention recently, not only from linguists but also from translators and IT professionals involved in processing natural language. Descriptions and propositions that take into account

---

24 "[...] una lexicografía basada fundamentalmente en la comunicación y que parte del valor intrínseco del vocabulario en el proceso de la comunicación, de los modos de uso y de las situaciones de uso de una unidad léxica dentro de una colectividad lingüística."

morphological, syntactic, semantic, and pragmatic aspects have arisen in order to identify and to recognize such compositions as well as to understand their limitations and structural patterns. In this context, there is a consensus that phraseology encompasses a great variety of word combinations, from collocations to proverbs (Ortiz Alvarez 2013).

Brazil has advanced in its study of the complex and broad field of phraseology. This denomination is prevalent in the country, even though it does not erase the formal and conceptual differences that these units involve. Following the tradition of other countries, Brazil has consolidated a field of studies that focuses on groups of words that come together in a more or less fixed way and whose meaning is understood as a set of its components, regardless of the understanding of each of its terms separately. These brief references reflect the wider body of factors involved in this type of structure.

In light of this comprehensive conception, the Brazilian panorama of phraseological research has expanded, especially since the beginning of the 21st century. Conferences, doctoral thesis and dissertations in universities have all helped to establish an investigative framework that has advanced in several directions, based on the interest of linguists, translators and computer scientists involved with natural language processing.

Ortiz Alvarez (2013) presents the state-of-the-art of phraseological studies in Brazil, making reference to the first authors who organized phraseological works, thesis and dissertations, citing research projects and a list of books, articles, monographs and dissertations. Most of these refer to the phraseology of the commonly spoken language, but there are also works on specialized phraseology which is part of professional communication. This phraseological typology is 'a representative structure of a conceptual nod of different thematic areas, especially when it includes a term in its composition' (Krieger/Finatto 2004, 85).[25]

Based on data collected in universities, specialized journals and the GTLex site, Bevilacqua (2020) identified two major trends in relation to specialized phraseologies. The first group is characterized as collocations, identified by morphosyntactic patterns. It has two basic elements: a term, usually a noun, considered the core, and an element that is usually an adjective or a verb. Thus, next to the term *heat* in texts about the environment, the collocations *absorção de calor* 'heat absorption' and *calor absorvido* 'absorbed heat' can be found (Bevilacqua 2004).

The second group comprises longer phraseologies or discursive forms considered formulas; in general, these are a characteristic of the textual genre to which they belong, such as the expression *Revoguem-se as disposições em contrário* 'all provisions to the contrary are hereby revoked', common in legal texts. Recurrent use and meaning in professional communication renders them conventional, functioning as discur-

25 "Uma estrutura representativa de um nódulo conceitual das diferentes áreas temáticas, sobretudo quando inclui um termo em sua composição."

sive stereotypes. Another typical case is the use of the verb *fazer* 'to make' in medical jargon, where it is used instead of *ter* 'to have' in cases like *o paciente fez febre* literally 'the patient made fever'.

The phraseological units of both groups have in common the fact that they are pluriverbal, but are distinguished by their length: in collocations, they are formed mainly by two elements, while in the discursive forms, they can extend to the length of a sentence.

Phraseological units of whatever kind are also of great interest to translators since they correlate with the lexicon and phraseological structures in many languages. In Brazil, it is common to refer to a heavy rain saying *está chovendo canivete* literally 'it is raining pocketknife', the English equivalent being *it's raining cats and dogs*. In translation, proverbs also become the object of semantically corresponding correlations. In these cases, the semantic equivalence principle is opposed to that of literal translation.

A variety of practical results from academic research on specialized phraseology is available. Some of these can be consulted online, as with the Termisul project. This group developed the Environment Law Lexical Combinatorics Database CLES[26] with Portuguese entries and their German, Spanish, French, English and Italian equivalents. The database records both of the aforementioned types of units: collocations, e.g., *coleção de óleos* 'oil collection' and formulae, such as *Esta lei entra em vigor na data de sua publicação* 'This law takes effect on the date of its publication'.

These are only some basic references of phraseology studies in Brazil that have gained increased visibility recently. Regardless of the categorizations and theoretical frameworks adopted, such studies show the connection of the common components between simple words, composed words and phraseology. These forms also serve to bring into focus the subtle distinction between lexical words per se and phraseological structures, when it comes to naming, establishing references, and conveying meaning, which are the essential roles of the lexicon in languages.

6 Final remarks

The previous sections have provided an overview of current research into the BPL, noting its heterogeneity and the different approaches taken in terms of description and analysis, including contemporary corpus-based methods. The chapter has also provided an account of the most notable lexicographic works in Brazil and their development throughout the 20th and 21st centuries, in which digitalization has played a major role over the most recent decades.[27]

26 http://www.ufrgs.br/termisul/cles/, last accessed 18.02.2022.

27 Translated by Cláudia Flores Pereira, revised by Rosalia Neumann Garcia (Lecttura Traduções) and John Barlow.

7 References

Almeida, Gladis Maria de Barcellos, et al. (2013), *Vocabulário Ortográfico Comum (VOC): constituição de uma base lexical para a língua portuguesa*, Estudos Linguísticos 42/1, 204–215.

Alves, Ieda Maria (2001), *Glossário de termos neológicos da economia*, São Paulo, USP.

Alves, Ieda Maria (ed.) (2010), *Neologia e neologismos em diferentes perspectivas*, São Paulo, Paulistana.

Berber Sardinha, Tony (ed.) (s.a.), *Corpus brasilileiro*, São Paulo, PUCSP, http://corpusbrasileiro.pucsp.br/cb/Inicial.html, last accessed 18.02.2022.

Bevilacqua, Cleci Regina (2004), *Unidades fraseológicas especializadas: aspectos semânticos*, in: Maria da Graça Krieger/Aparecida Negri Isquerdo (edd.), *Ciências do Léxico: lexicologia, lexicografia, terminologia*, vol. 2, Campo Grande, UFMS/PPG-Letras UFRGS, 275–287.

Bevilacqua, Cleci Regina (2020), *Fraseologia especializada: panorama das pesquisas realizadas no Brasil*, in: Suzete Silva (ed.), *Fraseologia e Cia: entabulando diálogos reflexivos*, vol. 2, Campinas, Pontes, 41–65.

Bevilacqua, Cleci Regina/Reuillard, Patrícia Chittoni Ramos (2010), *Acervo Termisul: linguagem legal, normativa e científica*, in: Aparecida Negri Isquerdo/Maria José Bocorny Finatto (edd.), *As Ciências do Léxico: lexicologia, lexicografia, terminologia*, vol. 4, Campo Grande/Porto Alegre, UFMS/UFRGS, 433–448.

Biderman, Maria Tereza Camargo ([2]2001), *Teoria Linguística: teoria lexical e linguística computacional*, São Paulo, Martins Fontes.

Borba, Francisco da Silva (2003), *Organização de dicionários: uma introdução à lexicografia*, São Paulo, UNESP.

Cabré, Maria Teresa (1993), *La terminología: teoría, metodología, aplicaciones*, Barcelona, Antártida/Empúries.

Cardoso, Suzana Alice Marcelino da Silva/Mota, Jacyra Andrade (2012), *Projeto Atlas Linguístico do Brasil: antecedentes e estágio atual*, Alfa: Revista de Linguística 56, 855–870.

Dick, Maria Vicentina de Paula do Amaral (2001), *O sistema onomástico: bases lexicais e terminológicas, produção e frequência*, in: Aparecida Negri Isquerdo/Ana Maria Pinto Pires Oliveira (edd.), *As Ciências do Léxico: lexicologia, lexicografia, terminologia*, 2ª edição, Campo Grande, UFMS, 79–90.

Finatto, Maria José Bocorny/Zilio, Leonardo (edd.), *Textos e Termos por Lothar Hoffmann*, Porto Alegre, Palotti.

Haensch, Günther/Wolf, Lothar (1982), *Introducción*, in: Günther Haensch et al. (edd.), *La lexicografía: de la lingüística teórica a la lexicografía práctica*, Madrid, Gredos, 11–20.

Isquerdo, Aparecida Negri/Seabra, Maria Cândida Costa (2010), *Apontamentos sobre hidronímia e hidroponímia na fronteira entre Mato Grosso do Sul e Minas Gerais*, in: Aparecida Negri Isquerdo/Lídia Almeida Barros (edd.), *As Ciências do Léxico: lexicologia, lexicografia, terminologia*, vol. 5, Campo Grande, UFMS, 79–98.

Krieger, Maria da Graça, (2008), *Políticas públicas e dicionários para escola: o Programa Nacional do Livro Didático e seu impacto sobre a lexicografia didática*, Cadernos de Tradução 2/18, 235–252.

Krieger, Maria da Graça (2012), *O dicionário em sala de aula: guia de estudos e exercícios*, Rio de Janeiro, Lexikon.

Krieger, Maria da Graça (2014), *Heterogeneidade e dinamismo do léxico: impactos sobre a lexicografia*, Confluência 46, 323–334.

Krieger, Maria da Graça (2019), *Terminologia Linguístico-Textual (TLT): princípios teóricos e metodológicos*, in: María Laura Perassi/Martín Tapia Kwiecien (edd.), *Palabras como puentes: estudios lexicológicos, lexicográficos y terminológicos desde el Cono Sur*, Córdoba, Buena Vista, 117–127.

Krieger, Maria da Graça (2020), *Lexicografia: a dicionarização do léxico*, in: Renato Rodrigues-Pereira/Daniela de Souza Silva Costa (edd.), *Estudos em lexicografia: aspectos teóricos e práticos*, Campinas, Pontes, 13–31.
Krieger, Maria da Graça/Finatto, Maria José Bocorny (2004), *Introdução à terminologia: teoria e prática*, São Paulo, Contexto.
Krieger, Maria da Graça et al. (2006), *O século XX, cenário dos dicionários fundadores da lexicografia brasileira: relações com a identidade do Português do Brasil*, Alfa: Revista de Linguística 50/2, 173–187.
Lara, Luis Fernando (2006), *Curso de lexicología*, México D. F., El Colegio de México.
Ortiz Alvarez, Maria Luiza (ed.) (2013), *Tendências atuais na pesquisa descritiva e aplicada em fraseologia e paremiologia*, São Paulo, Pontes.
Petter, Margarida Maria Taddoni (2006), *As línguas africanas no Brasil*, in: Suzana Alice Marcelino Cardoso/Jacyra Andrade Mota/Rosa Virgínia Mattos e Silva (edd.), *Quinhentos anos de história linguística do Brasil*, vol. 1, Salvador, Secretaria de Cultura e Turismo do Estado da Bahia, 117–142.
Rey, Alain (1977), *Le lexique: images et modèles. Du dictionnaire à la lexicologie*, Paris, Armand Colin.
Rodrigues, Aryon Dall'Igna (2013), *Línguas indígenas brasileiras*, Brasília, Laboratório de Línguas Indígenas, http://www.letras.ufmg.br/lali/PDF/L%C3 % ADnguas_indigenas_brasiliras_RODRIGUES,Aryon_Dall'Igna.pdf, last accessed 28.02.2022.
Senado Federal (2016), *Constituição da República Federativa do Brasil*, Brasília, Secretaria de Editoração e Publicações.
Timbane, Alexandre António/Coelho, Dayanny Marins (2018), *Os neologismos e a ampliação lexical nas redes sociais*, RELACult 4/1, 1–21, https://periodicos.claec.org/index.php/relacult/article/view/565, last accessed 18.02.2022.

Dictionaries

Academia Brasileira de Letras ([6]2021), *Vocabulário Ortográfico da Língua Portuguesa (VOLP)*, Rio de Janeiro, Academia, https://www.academia.org.br/nossa-lingua/busca-no-vocabulario, last accessed 18.02.2022.
Barroso, Hildebrando, et al. (1938), *Pequeno Dicionário Brasileiro da Língua Portuguesa*, Rio de Janeiro/São Paulo, Civilização Brasileira.
Biderman, Maria Tereza Camargo/Murakawa, Clotilde de Almeida Azevedo (2021), *Dicionário Histórico do Português do Brasil: séculos XVI, XVII e XVIII*, Araraquara, FCL-UNESP, http://dicionarios.fclar.unesp.br, last accessed 18.02.22.
Borba, Francisco da Silva (2002), *Dicionário de Usos do Português do Brasil*, São Paulo, Ática.
Borba, Francisco da Silva (2004), *Dicionário UNESP do português contemporâneo*, São Paulo, UNESP.
Caldas Aulete, Francisco Julio (1958), *Dicionário Contemporâneo da Língua Portuguesa*, Rio de Janeiro, Delta.
Ferreira, Aurélio Buarque de Holanda ([2]1986, [1]1975), *Novo Dicionário Aurélio da Língua Portuguesa*, Rio de Janeiro, Nova Fronteira.
Figueiredo, Cândido de ([4]1926), *Novo Dicionário da Língua Portuguesa*, Lisboa, Arthur Brandão.
Figueiredo, Cândido de ([14]1949), *Dicionário da Língua Portuguesa*, Lisboa/Rio de Janeiro, Bertrand/Jackson.
Flores, Valdir do Nascimento, et al. (2009), *Dicionário de Linguística da Enunciação*, São Paulo, Contexto.
Freire, Laudelino (1939–1944), *Grande e Novíssimo Dicionário da Língua Portuguesa*, Rio de Janeiro, A Noite.

Houaiss, Antônio (2001), *Dicionário Houaiss da Língua Portuguesa*, Rio de Janeiro, Objetiva.
Instituto Internacional da Língua Portuguesa (s.a.), *Vocabulário Ortográfico Comum da Língua Portuguesa (VOLP)*, Praia, https://voc.cplp.org, last accessed 18.02.2022.
Krieger, Maria da Graça et al. (1998), *Dicionário de direito ambiental: terminologia das leis do meio ambiente*, Porto Alegre/Brasília, Ed. da UFRGS/Procuradoria Geral da República.
Nascentes, Antenor (1961–1967), *Dicionário da Língua Portuguesa*, Rio de Janeiro, Academia Brasileira de Letras/Bloch Editores.

Rodolfo Ilari and Renato Miguel Basso

16 On the history of semantic studies in Brazil

Abstract: The term "semantics" has been used with various meanings and purposes since the 19th century, when several grammarians used semantic arguments in support of their theses (Guimarães 2004). During the last 50 years, many research trends in Brazil have defined themselves as "semantic", while arguing for competing theoretical hypotheses and methodological designs. Other research trends have emphatically refused the semantic label or have decided to append qualifiers such as "cognitive," "argumentative," or "formal" to it. Given this background, we adopt a four-step strategy in this survey. In our first step, we recall the motivational wanderings that led to the earliest Brazilian research groups adopting the name "Semantics". In a second step, we list these groups and try to evaluate their achievements, claiming that they attained uneven stages of development. This picture ignores many works that were not assumed to have been doing semantics but helped to uncover several important semantic features of Brazilian Portuguese (BP); therefore, our next section will be about these authors who did semantics *sans le savoir*. Finally, we will discuss a possible semantic agenda for BP, exemplifying both with traditional topics and topics that have come to the fore more recently.

Keywords: aspect, bare singular nouns, gerundismo, gerund vs. present participle, nouns vs. adjectives, passive voice, semantic schools in Brazil, semasiological and compositional approaches to aspect, *se* as a passive operator in Brazilian Portuguese, *ser* vs. *estar*

1 Historical background

1.1 The great turns

The understanding of semantics that prevailed in Brazil until the 1960s is exemplified by Bueno (1965/1947). Given its title (literally 'A treatise on Brazilian semantics'), this book should provide a definition of semantics as well as a semantic incursion into the BP of its time; we find instead long digressions on language arbitrariness and origins, its psychological or social nature, and other subjects that would not be considered semantics by the contemporary definition. Most of the information regarding BP evokes old etymological issues (such as the changes in meaning that turned *veado* < lat. *VENATUM* 'hunted' into a synonym for *cervo* 'deer', and *gado* < Lat. *GANATUM* 'won' into a farming word 'cattle', or the native meaning of toponyms like *Butantã* (name of a São Paulo neighborhood, 'hard ground' in Tupi). Bueno (1965/1947) also contains chapters on lexical borrowings and slang ("gíria"). Instead of an analysis of these areas of

https://doi.org/10.1515/9783110405958-017

BP, however, it provides advice on the best way to cope with them. As for borrowings, the recommended policy is tolerance, arguing that lexical borrowings have always existed; as for slang, it is described as being limited to Rio de Janeiro and to criminals—a mistaken and biased opinion. All these features reflect the historicist and prescriptive manner of viewing language that prevailed in Brazil until the 1960s and suggest that semantics was at that time a kind of no man's land where one could dump any subjects that were lacking adequate treatment.

In the 1960s, the blueprint of linguistics was redrawn several times. This was due to (i) the advent of Structuralism, which redirected Brazilian researchers towards synchrony and branded etymological issues as obsolete and (ii) to the impact of Generative Grammar and its program aimed at explicitly representing linguistic competence. As could be expected, the new ideas affected semantics, but changes in this area were somewhat different than in other linguistic fields and countered what would have been reasonable expectations; for instance, looking at the formal semantics currently practiced in Brazil, one could imagine, wrongly, that it came along with the generative grammar of the 70s. The success of Generative Grammar turned North American linguistics into the main reference for Brazilian linguistics,[1] wiping French style structuralism off the map, but Semantics and Generative Syntax evolved independently for two decades, and Davidson's prophecy (1967a) that Formal Semantics and Generative Syntax would eventually join forces was fulfilled in Brazil only at the end of the century. A kind of agreement was then attained that semantics must interpret the representations prompted by syntax, or, in other words, that the lower boundary of semantics is given by syntax. This agreement still prevails, despite one important questioning.[2] On the contrary, the "upper" boundary, where semantics borders Pragmatics and Discourse Analysis, remains controversial in Brazil (Negri 2006).

By the end of the 20th century, Brazilian universities came into contact with several trends opposing Generative Linguistics. Among these were (1) grammaticalization (see Votre 1994);[3] (2) the linguistics promoted in the eastern U. S. by authors like Fillmore, Lakoff, Kay, Rosch, and Sweetser (described in Franchi 2012/1982 as the "Berkeley Group"); and (3) Functional Grammars, developed in the Netherlands by Dik and

1 There had been earlier contacts with North America—the most important Brazilian romanist, Theodoro Henrique Maurer Jr., lived in Yale, where Leonard Bloomfield was his MA adviser. Mattoso Câmara Jr. also lectured in Austin in the 1960s.

2 Castilho (2007, 2010) claims that structuralism and Generative Grammar are mistaken in their assumption that the output of one component of linguistic theory should be the input of a superordinate component. He conceives language as a complex adaptive system where one recognizes four independent components (syntax, lexicon, semantics, and pragmatics) linked by a socio-cognitive device. The relationship among these components is established by cognitive exchanges, not by logical priority.

3 For older mentions to grammaticalization, see Câmara Jr. (1975).

Hengeveld and in the U. S. by Langacker and Talmy. All these theories had influence in Brazil.

Grammaticalization entails semantic changes, the most common being the effects of "bleaching", by which referential items develop abstract meanings typical of grammatical morphemes. The most quoted example of grammaticalization in Portuguese is the development of the 2nd person pronoun *você(s)*, in which bleaching caused the oblivion that the departing expression *vossa mercê* named the only individual capable of mercy, namely the king. Many remarks of this kind were made, but the gains for semantics were poor because in Brazilian linguistics, grammaticalization research has focused on a small number of unrelated forms.[4] The only important exception to this tendency concerns Portuguese conjunctions. In Old Portuguese, many conjunctions inherited from Latin (*ca*, *pero* e *porém* < lat. QUIA, PER HOC, PER + INDE, *pois* < lat. POST, etc.) were replaced by new versions where the "universal connective" *que* follows a preposition, participle, or adverb. That's the origin of present day conjunctions *porque*, *depois que*, *desde que*, *apesar de que*, *posto que*, *dado que*, *já que*, *mesmo que*, *a menos que*, *só que*, etc. (Barreto 1999; Poggio 2002). All the words that precede *que* in these compounds have gone through subtle changes of meaning. For instance, the *só* that intervenes in *só que* means opposition rather than uniqueness (Longhin, 2003).

Dutch functionalism has been studied in some universities of Southwestern Brazil, and Talmy and Langacker influenced the "Abordagem Multissistêmica da Língua" argued for by Castilho (2009; 2010).

As for the "Berkeley ideas", we recall that they consisted, among others, in conceiving categorization as a prototype-based activity (Rosch 1973; Taylor 1989), in electing Construction as the basic concept of syntax (Fillmore et al. 1988; Goldberg 1995), and in believing that humans understand any aspect of reality taking as models previous and bodily experiences (Fauconnier 1994; Lakoff 1982). Not only are these ideas nowadays the hallmark of Brazilian "Prototype Semantics" (see below), but they also proved useful in other linguistic areas: An example of this is Camacho (2011), where it is argued that (1) words derived by nominalization constitute a transitional word class between the prototypical classes of nouns and verbs (as they have noun distribution but also have arguments); (2) the whole array of arguments of a cognate verb is operative in the interpretation of these words, even when some arguments are omitted for communicative or exactness reasons.

The first of these claims is an application of prototypes to grammar, and this application helps understand one of the most frequent usages of *de*—introducing the arguments of a nominalization. The second claim touches a point that returns whenever comparisons between BP and EP are made: what contents can be omitted in running

4 For a list of grammaticalization facts in Portuguese, see Gonçalves et al. (2007).

conversations without loss of information (in syntax, this problem concerns the *objeto direto nulo* 'null direct object', exemplified by sentences such as *Comprei uma revista e li no avião* [literally: 'I bought a magazine and read on the plane', where *li* is interpreted as 'I read this particular magazine' in BP]). Last but not least, by dealing with words created by nominalizations with an eye on their arguments, Camacho (2011) rejoins a tradition that is particularly strong in Brazilian Semantics and Lexicology, and exemplifies a host of works departing from non-semantic motivations that have achieved significant semantic contributions.

1.2 Other actions

Structuralism and Generative Grammar were effective in updating Brazilian Linguistics to the world vanguards, but in order to explain the emergence of a semantic tradition in Brazil, one must also consider some facts that are often neglected. In what follows, we recall the presence of semantic titles in the publishing market (1.2.1), the emergence of corpus linguistics (1.2.2), and scholarly contact and exchanges that helped circulating seminal concepts (1.2.3).

1.2.1 Semantics in the publishing euphory

In the late sixties, Linguistics was recognized as the leading science for humanities, and it gained an unprecedented presence in the Brazilian publishing market. Titles that were released were mostly, though not exclusively, translations from English. Two significant translations concerning semantics were Bonomi/Usberti (1971) and Lyons (1976/1970, for Bierwisch's chapter on componential semantics). Lobato (1977) contained among others the translations of Weinreich (1966) and Katz/Fodor (1963). Semantic writings by Bloomfield, Halliday, and Davidson became available for the first time in Portuguese in Dascal (ed., 1982c). All these translations gave some popularity to their original authors—when John Lyons published *Semantics* (1977), he was already well known in Brazil, which explains the success of this book that quickly became a reference for an entire generation.

In the 1970s, books and chapters on semantics for undergraduate students majoring in Portuguese came into vogue. Only a few of these introductions equal their foreign equivalents in comprehensiveness, but new titles of the genre continued to be published. As these works usually elaborate on what distinguishes semantics and pragmatics, they may have helped to keep alive this distinction on which remarkable works had been written abroad (e.g., Wilson 1975a; 1975b). It is fair to say that the situation regarding Semantics textbooks in Brazil has since improved: there are some books written directly in Portuguese, dealing with Portuguese examples and problems rather than translated works with examples adapted from other languages (e.g., Can-

çado 2012; Ilari 2001; 2002; Ilari/Geraldi 1985; Oliveira 2001; Ferreira 2019; Gomes/Mendes 2018).[5]

1.2.2 The emergence of Corpus Linguistics

Some important research programs based on corpus linguistics were launched around 1970. The most important one was the *Projeto de Estudo da Norma Urbana Linguística Culta*, designed to collect samples of the speech used by learned inhabitants of the most important state capitals at the time. Usually referred to by the shorthand NURC, this nationwide project was the most outstanding project in Brazilian Linguistics for three decades and was carried forward by other nationwide projects, such as the *Projeto da Gramática do Português Culto Falado no Brasil* and the *Projeto de História do Português Brasileiro* (PHPB).

The NURC was expected to gather information on the lexicon of its time; in order to do so a specific subject was assigned to each interview, and interviewers were asked to elicit from their subjects not only the report of past experiences, but also the vocabulary routinely associated with analogous experiences.[6] The expectation of producing an inventory of BP lexicon was partly unfulfilled, but, as an unexpected result, the NURC spread the idea that the extant monolingual dictionaries (such as Aurélio Buarque de Holanda et al.'s dictionary, with various editions starting in 1938) were precariously connected with the living language. Consequently, the NURC project helped to legitimatize a new kind of dictionaries, based on large corpora of written language and represented by Borba (1990; 2002). By including only attested words with a frequency above a certain threshold, these dictionaries were far more reliable as a representation of the contemporary Brazilian lexicon. One of Borba's dictionaries contains only verbs and describes their meanings and the arrays of arguments found in corpora. Apart from terminology, Cançado et al. (2013) covers the same topics.

---

5 It should also be noted the many Brazilian researchers are now written materials aimed at students and teachers from more basic levels of education, such as http://anamuller.fflch.usp.br/sites/anamuller.fflch.usp.br/files/inline-files/Semântica%20na%20escola.pdf, last accessed 18.02.2022 and https://negr.paginas.ufsc.br/files/2020/05/Artefatos-em-gramática.pdf, last accessed 18.02.2022.

6 For instance, in a 1968 interview recorded in São Paulo on "Movies, television, radio, and theater", we find these conversation turns: "*Entrevistador: Antes de entrar no cinema a senhora... [...] o que a senhora faz?* / *Informante: Bom, adquiro o bilhete para entrar*" ("Interviewer: Before entering the theater, you... [...] what do you do? / Informant: Well, I buy the ticket to be allowed in"). In this question, the interviewer asks something apparently obvious in order to obtain from the informant the word specifically used in that time for buying movie tickets, i.e. "*bilhete*". By asking some hundreds of questions of this kind, the researchers of NURC expected to gather and testify a great amount of semantic and syntactic information (meaning, government, collocations etc.) on words that had never received any attention from BP researchers. Unfortunately, only a fraction of the collected materials was processed during the subsequent years.

During the last two decades of the past century and thereafter, the assemblage of corpora and corpus linguistics as a set of methods and technologies has only increased. Among the corpora that are currently used in Brazil are the CETENFolha, C-Oral, PEUL, PROHPOR, Tycho Brahe, and VARSUL. Corpus Linguistics has reinforced the commitment to using recorded materials as evidence for any linguistic hypothesis, so much so that the reference to attested sources has become essential for many linguists. As a single example, when Longhin (2012) claims that the word *pois* denotes different manners of linking sentences in texts belonging to different discursive traditions, she gives as evidence several lists of occurrences found in letters written during the 17th century by civil servants of the Portuguese colony as well as other occurrences from the 19th century by newspaper readers.

Finally, the manual compilation of the first *corpora* foretells the preparation of dictionaries for specific purposes and the current production of computer-aided tools in some Brazilian universities, as will be outlined below.

1.2.3 Contacts, exchanges and new ideas

The first of these three words—contacts—will be taken in a broad sense, including not only Brazilians who received some training in the northern hemisphere, but also foreign linguists who came to Brazil for conferences as well as for the first courses in semantics ever offered.[7]

Sending scholarship students abroad became a regular policy by the end of the sixties, when the funding agencies (mainly the *Fundação de Amparo à Pesquisa do Estado de São Paulo*, founded in 1960) started funding groups of young researchers for training abroad. One of these actions led to the foundation of the first Linguistics Department in Brazil (Campinas, UNICAMP, 1977), which played an important role in the development of linguistic and semantic thinking.

As for the visits of foreign scholars and their importance for opening a new semantic trend, they were very frequent, but their effects were uneven. In some cases, the presence of foreign researchers led to the formation of full-fledged semantic schools, through a serious discussion of the new theoretical concepts and research methods. This was the case for "Argumentative Semantics". In many other cases, however, the imported theories were almost entirely abandoned after an enthusiastic welcome. For instance, Greimas' structural semantics motivated the foundation of "supporter clubs" in some towns in the region of São Paulo[8] but did not proceed past

7 Two names deserving of mention in this context are A. J. Greimas and O. Ducrot, who both lectured on their own theories.

8 Ignácio Assis Silva and others organized several "Greimas courses" in inland São Paulo and founded a journal ("BACAB") dedicated to Greimas' semantics and semiotics.

the stage of contact with the original model, leaving only a very tiny amount of descriptive work.

In a still broader sense "contacts and exchanges" includes the circulation of key ideas and works. Two books with such a role were Reichenbach (1947) and Vendler (1967) which became known in Brazil around 1970 and gave new guidelines to the study of verb aspect. This subject had been addressed by Castilho (1967), starting from tenses and searching out possible meanings and vice-versa (this approach was adopted in ensuing works such as Travaglia, 1981). When Reichenbach (1947) was "discovered", its application helped separate temporal and aspectual options, settling misunderstandings caused by the ambiguous terminology used by grammarians. For instance, it became clear that Imperfect Indicative and Perfect Indicative, two different tenses, both denote the same time location; or that the main device for locating an ongoing action at speech time is not the Simple Present, but the periphrasis containing *estar* + the gerund (cp. *O gelo derrete a zero graus* and *O gelo da calota polar está derretendo*, 'Ice melts at 0° Celsius', 'The ice of polar caps is melting'). This opened a trail along which some old challenges were re-examined. One such challenge consisted in explaining what distinguishes two Portuguese past tenses—the simple (as in *Vi o Zé*, 'I saw Joe') and the compound, PCP (as in *Tenho visto o Zé* 'I have seen Joe'). Contrasting with its Romance analogues, the latter conveys an idea of repetition (i.e., the event of seeing Joe has already occurred several times and is expected to occur even more times in the future), and it is not always obvious what in the event is subject to repetition. For instance, in the sentence *João tem pulado deste trampolim* 'John has jumped/has been jumping from this springboard', the repeated event is John's jumping from a certain springboard. In *Muitos têm pulado desta ponte para a morte* 'Many (suicidal) people have jumped/have been jumping to death from this bridge', however, what is repeated is not collective suicides, as the sentence can only be true if different people commit individual suicides on different occasions. From examples like these, it becomes clear that PCP's interactions with the quantification of verb arguments is rather convoluted[9] and that its explanation requires a fairly rich notion of event. This requirement led naturally to Davidson's (1967b) writings on the logical form of action sentences, which became familiar to Brazilian linguists at the end of the 20th century (Ilari 1998a; Guimarães 2002).

The famous fourfold distinction of action classes into *accomplishments*, *achievements*, *states* and *activities* was first revealed to Brazilian linguists by a copy of Vendler (1967) that the philosopher Andrea Bonomi had brought along in 1972. In the simplest cases, this classification confirmed the intuition already present in Castilho (1967) that there is more than time location to the semantics of tense. But simple cases were soon overcome, and a fundamental step was then taken—as soon as it was un-

9 On the Portuguese compound past, see Viana (1901), Boléo (1937), Ilari (1998b), and Ilari/Basso (2014, 190–195).

derstood that questions about *Aktionsart* should be asked not about the verb but about the verb and the sentence as a whole, a new, compositional program emerged, aimed at understanding how the meaning of a sentence is related to the meaning of its components. This new program is still producing results, mainly works on the completeness, quantification, and pluralization of events. By adopting a compositional view, many authors were led to assign words a more articulated semantic structure than the one which surfaces from syntax. For instance a BP sentence like *Esta casa foi alugada por dois anos* (literally 'This house was rented for two years') is true in two very different situations; (1) when a lease was convened upon with a two-year duration, and (2) when (independently from any lease) the house was occupied for two years by a tenant. This is possible because the adjunct *por dois anos* can measure two different durations predicted in the lexical entry of *alugar*. This type of analysis is still carried on by many scholars and is one of the hallmarks of Lexical Semantics.[10]

2 Semantics in Brazil, a variety of schools

Summing up what has been seen so far, the history of Semantics in Brazil was a history of imports or rather of import substitutions. Theories and ideas stemming from a variety of sources contributed to the creation of a research agenda focused mainly on synchronic issues and potentially capable of bringing forth descriptive works. In what follows, we describe this landscape using a classification of "semantic schools" found in Ferrarezi Jr./Basso (2013).

The authors distinguish nine semantic schools by prefacing Semantics itself with nine qualifiers: 1. Argumentative, 2. Cognitive, 3. Computational, 4. Cultural, 5. Enunciation, 6. Prototype, 7. Experimental, 8. Formal, and 9. Lexical. This number suggests that the current study of Semantics in Brazil is fragmented, but this conclusion is only partially true. It is a fact that the times when a single theory seemed hegemonic (as was the case for Greimas' Structural Semantics as well as Ducrot's Argumentative Semantics) are now behind us. But the semantic schools listed in Ferrarezi Jr./Basso (2013) are less different from each other than it might seem. For example, Prototype, Cognitive, and, to some extent, Experimental Semantics are tributary of the ideas of the "Berkeley group" and can be described jointly; the use of computers no longer seems to be a valid criterion to distinguish one school from another; Enunciation Semantics appears to be a development of Argumentative Semantics; and finally, both Formal and Prototype Semantics attribute a representative function to human language, as opposed to Enunciation and Argumentative Semantics, which conceive lan-

10 For an overview on the semantics of tense and aspect in BP and European Portuguese (EP), cf. Ilari/Oliveira/Basso (2016), and for an overview on lexical semantic studies in BP and EP, cf. Cançado/Anabela (2016).

guage as a stage for struggles between interlocutors. Considering these different intersections, we have organized what follows in fewer sections.

2.1 Argumentative Semantics/Enunciation Semantics

Argumentative Semantics was introduced in Brazil in the 1970s by the French linguist Oswald Ducrot, and by outstanding followers, like Carlos Vogt, Eduardo Guimarães, and Wanderley Geraldi. Explained in books such as Ducrot (1972; 1980), Argumentative Semantics found supporters who, for an ideological bias, regarded language and dialogue as a place for engagement and manipulation (thus giving linguistics an interpersonal focus). Received enthusiastically, discussed at a level that allowed improvements of the original model,[11] and finally applied with productive results to BP,[12] this semantics fully completed the cycle of possible steps that can be taken by a novel theory. One of its descriptive concepts was particularly effective in describing BP: argumentative scales. As the presence of scales is clearly denounced by the use of certain conjunctions (starting from *mas* 'but'), much work was dedicated to revealing the argumentative strategies covered by conjunctions (Guimarães 1987).

Argumentative Semantics also concerned itself with the phenomenon of presupposition and explained it in a manner different than philosophers such as Frege and Strawson had done. The study of presupposition by Argumentative Semantics pointed to the opposition between presuppositions and implied information, a subject that would soon become a research field in its own right. The argumentative explanations of implicit contents constituted a robust alternative to the Gricean concept of conversational implicature, which was the preferred subject for Brazilian philosophers of language during the seventies and eighties and was regarded for some time as semantic fact (later on, implicatures would be seen as the pragmatic phenomenon *par excellence*).[13]

Nowadays presupposition, implicit contents, and argumentative scales are old subjects in Ducrot's thought, who since the eighties has been increasingly interested in phenomena related to dialogism (such as *topoi*) and went on to propose the Semantic Blocks Theory (Carel 1992). These new developments had some circulation in Brazil, but they do not seem to have received important theoretical improvements nor to have led to any amount of descriptive analyses comparable to what resulted from the Argumentative Semantics of the 70s. In other words, of the three stages that we have

11 As is shown, among others, by Vogt (1977; 1980). Vogt (1980) contains a Portuguese version of "De *magis* à *mais*, une hipothèse sémantique", published originally in French (Ducrot/Vogt, 1979).

12 Cf. Figueira (1974), Geraldi (1978), Zamboni (1978), and Taffarello (1979).

13 Some cases of implicature were analyzed in Dascal (1982a); also see Dascal (2006b). For an updated introduction to the formal pragmatic theory of implicatures written in Brazil, cf. Oliveira/Basso (2014).

distinguished regarding theoretical imports — (i) assimilation and circulation of a model, (ii) critical discussion, and (iii) its use in the analyses of relevant aspects of BP—Ducrot's new ideas appear to be still in the first or second one. In other words, Ducrot's most recent theories have been circulated and debated in Brazil, but have not yet been applied to the analysis of relevant aspects of BP. In this sense, their influence on Brazilian semantics remains incomplete.

Being direct disciples of Ducrot, Brazilian proponents of Argumentative Semantics have always been receptive to the ideas of another French linguist, Émile Benveniste, and particularly to the ideas he presented in the chapter "L'homme dans la langue" in Benveniste (1966). These ideas form the core of what is known as Enunciation Semantics, practiced at different periods by Beth Brait, José Luiz Fiorin, Valdir Flores, Eduardo Guimarães, and Eleni Jacques Martin. These scholars added ideas from Bally, Guillaume, Jakobson, Greimas, Culioli, Bakhtin, and Hagège, and others to the core ideas presented by Benveniste. The influence of Benveniste is even more evident in the application to BP of the semantic categories that he called "the formal apparatus of enunciation" (Benveniste 1970), One such category is the pair "enunciative vs. enuncive" (Fiorin 2002), which helps to understand why a decontextualized analysis assigns two different meanings to sentences like (1):

(1) A próxima rua é a Rua Alagoas.
'The next street is Rua Alagoas.'

The enunciative interpretation refers immediately to the speech situation; on the other hand, the "enuncive" interpretation refers to something already uttered—a contrast that may be illustrated by placing (1) after (2) and (3):

(2) *Olha! Na próxima rua tem um estacionamento. Que rua é essa?/A próxima rua é a Rua Alagoas*
(Literally: 'Look! There is a parking lot in the next street. What street is this?/The next street is Rua Alagoas.')

(3) Siga em frente até o farol./A próxima rua é a Rua Alagoas.
(Literally: 'Go straight to the street lights./The next street is Rua Alagoas.')

Obviously, the enunciative/enuncive opposition establishes an important division among the words that mean temporal or spatial location: *hoje*, *ontem*, *agora*, *aqui*... 'today, yesterday, now, here...' are enunciative; *naquele dia*, *na véspera*, *então*, *lá*... 'on that day, on the previous day, then, there...' are enuncive; *importar* and *exportar* 'to import', 'to export' can be both.

2.2 Cognitive Semantics/Prototype Semantics/Experimental Semantics

Among the Brazilian scholars who profess cognitive semantics, deserving of mention are the works of Maria Lúcia L. de Almeida, A. Bussons, K. Falcone, H. P. M. Feltes, Lilian Ferrari, Paula Lenz, A. C. P. Macedo, Mario E. Martelotta, Erick Miletta Martins, N. S. Miranda, A.C. Pelosi, Margarida Salomão, and Mara Sofia Zanotto. The cognitivism referred to here is associated with the so called Berkeley group. In Brazil, no one has represented these ideas more enthusiastically than Margarida Salomão and her students at the Universidade Federal de Juiz de Fora (RJ), who have struggled for a syntax based on constructions, a categorization based on prototypes, and a concept of cognition which is based on bodily experiences (cf. Lenz 2013).

Outside Brazil, prototype semantics is employed to explain the history of polysemic words and semantic fields (Geerarts 1997; 2010). In Portugal, Silva (2006) has shown, combining cognitivism with thoroughly historical research, that some meaning features which were first unimportant for the word *deixar* eventually became prototypical, giving this word its current meanings 'to abandon' and 'to allow'. In Brazil, the most popular application for this semantics was different. During the last several decades, pushing the claim that prototypical meaning reflect the speakers' interaction with the world, two researchers with storied and successful experience in Text linguistics, Ingedore Koch and Antônio Marcuschi, claimed that the reference of linguistic expressions is always the result of a negotiation (cf. Koch 2002); consequently, the traditional accounts of reference, hopelessly objectivistic, should be abandoned, and the notion of reference itself should be replaced by a new and more interactional one they called *referenciação* ('referencing'; with its ending *-ação* , this word itself declares that picking up a referent in the world is an action of speakers). Marcuschi and Koch's ideas have had a large application in the description of the thematic progression, the process by which the same character of a story is referred to by different descriptions as the story evolves. As should be expected, Koch and Marcuschi's ideas have found a strong opposition among researchers who believe that the meaning of linguistic expressions is relevant for reference and is a property of language.

A natural extension of Cognitive Semantics is the designing of models of the brain that explain linguistic phenomena—this is the main goal of Psycholinguistics according to Cunha Lima (2013). Psycholinguistics investigates language processing; that is, it looks at how words, sentences, and texts are processed and represented in the human brain. With respect to Semantics, Psycholinguistics tries to explain how we understand, store, and produce texts and linguistic fragments that make sense. This is done with the aid of large empirical and experimental collections of data as well as by conducting tests and experiments with subjects which constrain and empirically verify our language processing models. Brazil is the home to some important Psycholinguistic centers that are generally focused on syntactic and semantic phenomena—

among them, the federal universities of Minas Gerais and Rio de Janeiro as well as the Catholic University of Rio Grande do Sul.

2.3 Cultural Semantics

Celso Ferrarezi, Jr., the most well-known of the scholars who characterize their own research as Cultural Semantics, claims to follow Peter Burke, Roy Porter, Martin Jay, Tamar Katriel, Jessica Munns, and Gita Rajan, but he was also deeply influenced by Benjamin Whorf's ideas on determinism and relativism. It is no coincidence that this kind of Semantics emerged in Rondônia, an Amazonian state where contact with native cultures is still possible. The facts crucial to the semantic trend followed by this researcher concern the way language organizes experiences in different cultures and how the worldviews associated with native languages are affected by western civilization. An example from Ferrarezi Jr. (2013) typifies this interest, recalling an experience he shared with a member of the Oro Nao tribe: the two go fishing on a boat, and the native instructs the semanticist to catch small fish that will be used as bait. This task results in catching fish of different species that the native invariably calls *piabas*. Asked if the Oro Nao do not have names for these species, the native confirms that they do, but he has forgotten them. He explains that he is calling any small fish *piaba* because it means 'bait'. Afterwards, they catch bigger fish that are valuable in the region, and the native says their names in his native language. *Piaba* is not an Oro Nao word; it is a term that the natives borrowed from Portuguese. Forgetting the native names of these fish happened in parallel with a reorganization of the ichthyologic knowledge of their tribe, which adopted Western customs and so went on to distinguish between fish that have commercial value (which maintain their traditional designation) and those without it (which have lost their traditional designation). We know, of course, that changes of meaning of this kind have occurred by the hundreds in the history of Indigenous languages; they are worth recording while it is still possible, and they should be collected on a large scale. Unfortunately, for now, Cultural Semantics has not gone beyond a few isolated examples due to external circumstances such as decreased contact with natives, even in the Amazon.

2.4 Formal Semantics/Computational Semantics

Formal Semantics may have started in Brazil in the 1970s with the Logic and Language courses offered at the Universidade Estadual de Campinas thanks to the presence of several logicians and philosophers.[14] In these courses, works of scholars such as

14 Dascal (2006a), Oswaldo Porchat, and Lopes dos Santos (2008).

Frege, Russell, Strawson, Lewis, Davidson, and Montague were read for the first time in a Brazilian linguistics department. The students of these classes learned fundamental notions that would allow them to later teach these ideas in other universities. As a result, three generations of Formal Semantics in Brazil now exist, the last two being still at work.[15]

Taking for granted the compositional nature of sentence meaning, Brazilian formal semanticists have provided answers both for old issues and for new ones during the last 30 years: (i) Wachowicz (2003) has studied the progressive forms of verbs; (ii) Basso (2007) has discovered "detelicization" facts in sentences such as *João construiu a casa por um ano* (literally: 'John built the house for a year'); (iii) Gonçalves (2007) has investigated imperfectivity and its morphology; (iv) Molsing (2010) has studied the compound past tense in Portuguese; (v) Bertucci (2011) has explored aspectual auxiliaries; (vi) Ferreira (2004) has written on quantification and plurality of events. Assuming the truth-functionality of meaning, they have advanced explanations for at least the following facts of BP: (vii) cleft constructions (Menuzzi 2018; Teixeira/Menuzzi 2015); (viii) progressive form of verbs (Wachowicz 2003) ; (ix) cleft-constructions (Menuzzi 2012; Menuzzi/Rodrigues 2010); (x) comparative structures (Souza 2010); (xi) maximality, quantification, and distributivity (Gomes 2009); (x) genericity (Müller 2002; 2004); (xi) bare nominals (Oliveira 2010); (xii) countertdirectionality (i.e., the semantics of expressions such as *de volta pra casa* in sentences such as *João caminhou de volta pra casa*, 'John walked back home'—Zwarts/Basso 2016; Basso 2019a).

All of this research happened when, fulfilling Donald Davidson's prophecy, Formal Semantics and Generative Syntax converged into an integrated linguistic theory. This convergence is shown by handbooks such as Chierchia/McConnell-Ginet (1990) and Heim/Kratzer (1998), through which most Brazilian linguists have learned to view Semantics and Syntax as compatible research activities. This attitude is well represented in a book by Oliveira/Mioto (2011), which features chapters on syntactic matters—such as the syntax of comparative constructions (Souza 2011) and free relative sentences (Marchesan 2011)—side by side with chapters on semantic subjects such as the effect of adverbs ending in *-mente* on event structure in BP (Foltran 2011).[16]

Finally, it should be noted that thanks to the conditions created in the Brazilian context by its followers, Formal Semantics is now well institutionalized, with visibility in Brazil and abroad. A key factor for this dialogue were the exchanges that Brazilian formal semanticists have had with northern hemisphere universities, and, as a return, the organization in Brazil over the years of regular scientific meetings attended by many representatives of this school of thought.[17]

---

15 For a glimpse into the main themes of research investigated in Brazil in Formal Semantics, see *Cadernos de Estudos Linguísticos* (Campinas) number 52, Müller (2003), and Negri et al. (2004).

16 A more recent book, with the same attitude, is Oliveira et al. (2020).

17 Renowned foreign researchers have been present both in large events, such as the meetings of the *Associação Brasileira de Linguística*, and in specialized meetings, such as the "Workshop on Formal

Formal Semantics was the starting point for several researchers who now define themselves as "computational semanticists". The adjective "computational" in "Computational Semantics" identifies three branches of research where the computer plays a prominent role. One of these lines of research is the construction of algorithms in a computer language which could process natural language sentences to calculate their truth-conditions. The idea is, in this case, to simulate some fragment of our semantic competence through a language containing only controlled symbolic calculations (cf. Pagani 1998; Gonçalves/Pagani 2004). An entirely different field in which the use of computers has provided remarkable advances is the statistical treatment of large portions of the lexicon, and their equivalents in other languages, for the development of automatic translators (cf. Martins 2004). Yet another use of computers is being carried out at the *Núcleo Interinstitucional de Linguística Computacional* (NILC) based at the University of São Paulo (in São Carlos campus). This center has worked mainly with machine generated summaries and text simplification, machine assisted reading and writing, the creation of lexicons/dictionaries, corpus Linguistics, and automatic language analysis on various levels (parts of speech, syntax, semantics, and discourse).[18]

Completing this section, it is worth remembering that Formal Semantics has influenced other linguistic areas as well. For example, Ilari et al. (1996) classify adverbs by asking (1) which expressions the adverb operates on and (2) what kind of meaning modification is entailed by its presence; the first of these questions is an application of the logico-semantic notion of scope. The second is an attempt to understand why certain adverbs are not mere "predicates of predicates" in the logical sense of these terms.

2.5 Lexical Semantics

The term Lexical Semantics was used by Teresa Cristina Wachowicz (2013) to bring together researchers like herself, Márcia Cançado, Larissa Ciríaco, Elena Godoi, Marcos Lunguinho, and Rozana Naves. These researchers value some celebrated notions such as Fillmore's semantic cases (1966; 1977), Vendler's actional classes, Dowty's (1979) formalization of Vendler's insights using DO and CAUSE operators, and Dowty's pro-

Linguistics", an initiative of Ana L. Müller, maintained by the efforts of Marcelo Barra Ferreira, Ana Paula Quadros Gomes, Sergio Menuzzi, Esmeralda Negrão, Roberta Pires de Oliveira, et al. Among the visitors were Pascal Amsili, Claire Beyssade, Greg Carlson, Gennaro Chierchia, Carmen Dobrovie-Sorin, Jenny Doetjes, Irene Heim, Angelika Kratzer, Barbara Hall Partee, Susan Rothstein, Henriette de Swart, and Jost Zwarts. Many of these meetings resulted in important works that were accepted by journals such as *Letras* (UFPR), *Revista da Abralin*, and the *Journal of Portuguese Linguistics*.

18 For more on NILC (its history, teams, research topics, and products), see http://www.nilc.icmc.usp.br/nilc/index.php, last accessed 18.02.2022.

posal to explain semantic cases as entailments (1991). As could be predicted from these references, scholars working in this line of research focused on topics such as action classes and verbal aspect, as well as the implementation of verb arguments and other phenomena relating lexical and syntactic structures. These studies proposed several classifications for verbs in BP and showed that they are subject to particular types of alternations—one of which is the causative-inchoative alternation, shown in *O jogador rolou a bola no meio do campo* 'The player rolled the ball to the middle of the field' to *A bola rolou no meio do campo* 'The ball rolled to the middle of the field'.

Another important topic included in Lexical Semantics is the study of verbal periphrases and auxiliaries. Both EP and BP employ a large number of periphrases containing a conjugated verb plus a gerund, infinitive, or past participle. In addition to *ser* (as in passive voice) and *ter* (as in compound tenses), many other verbs such as *estar*, *andar*, *ir*, *começar a*, *continuar a*, *parar de*, *acabar de*, or *dar de* can fulfill this role; when they do, they lose their usual meaning 'to be, to walk, to begin, to continue, to stop, to finish, to give...' and denote subtle nuances of time, aspect, or modality in ways that are not (or no longer) possible in most Romance languages. For example:

(4) Como vai o senhor? Vou levando.
'How do you do, Sir? I'm carrying on.'

(5) Porque vocês estão preocupados com o José? Porque ultimamente ele deu de beber.
'Why are you so worried about Joe? Because he took up drinking.'

(6) Pedrinho, o que você está fazendo debaixo do sofá? Estou brincando de esconde-esconde.
'Pete, what are you doing under the sofa? I am playing peek-a-boo.'

(7) Por que a empregada está tão feliz? Porque hoje está tendo água.
'Why is the maid so happy? Because today there is no lack of water' (literally: 'there is having water')

One important topic is to establish whether these verbs are alike in distribution and make similar contributions to the meaning of sentences. Lobato (1970), Pontes (1971), and Ilari and Basso (2014) answered this question negatively, arguing that there are "degrees of auxiliarity" which can be measured by applying tests to each verb, such as the possibility of having different subjects for the conjugated and the unconjugated verb, or the possibility of denying them independently, or distinguishing auxiliaries from semi-auxiliaries. Notice that verbs that are synonymous when used as main verbs, such as *começar* (to start) and *iniciar* (to initiate) or *acabar* (to end) and *terminar* (to finish) are not equivalent as a part of a periphrasis, as shown in a pioneering work by Dascal (1982b).

Lexical semantics in Brazil has yielded some interesting results, but there is still a lack of consensus regarding its main research topics and methodology. The impression often is that there is a network of people working on similar topics rather than a particular semantic school.

3 Away from semantic schools: ideas of authors who did semantics *sans le savoir*

The different semantic trends that we have considered so far have brought to the fore distinct topics, and this could seem to be their main difference. In fact, the main contrast among semantic schools in Brazil, is associated with different ways of conceiving and ranking language functions (to organize humans' experience of the world, to convince others, to express a given culture, etc.). This plurality is welcome as far as it fuels the debate about language functions, but is also confusing as we try to compare. In an ideal situation, each of the competing trends could be associated to a program (theoretical principles and methodological directions illustrated by some crucial examples) and to a list of linguistic facts studied; the competing semantic schools could then be evaluated by comparing the results of their programs. But, as we have seen, this is not the state of affairs in Brazil, where many semantic schools never came up with strong descriptions of large portions of Portuguese.

Moreover, it would be unfair to identify Brazilian semantics solely with the activity of researchers that have adopted the term *semantics* as a qualifier. In the history of linguistic studies in Brazil, there exist works of high interest for semantics that were originally labelled by their authors as syntactic, stylistic, or even morphological and were written with no connection to the schools described in the preceding sections. These are works whose authors made semantics *sans le savoir*. Some of these works tackle old puzzles of Portuguese grammar, and it is important to retrieve the semantic ideas they elaborate on. In what follows, by way of example, we shall refer to writings on the following subjects: noun/adjective distinction, gerunds, the *ser*/*estar* opposition, and the passive operator *se*.

3.1 Nouns and adjectives

Distinguishing nouns and adjectives could be expected to be a mere morphological matter: after all, adjectives, unlike nouns, can take endings marking both masculine and feminine as well as singular and plural and have special forms for normal and superlative degrees. But this expectation fails because many adjectives are used only in normal degree (*justiça federal*/**justiça muito federal*, 'federal justice', '*a very federal justice') and many sequences (*velho japonês* or *conhecido socialista* literally, 'old Japa-

nese' and 'known socialist') admit alternative analyses where the noun is either the first or the second word (*velho japonês* 'a Japanese who is old' or 'an old man who is Japanese'). Lobato (1993) and Borges Neto (1991) investigated the factors involved in this choice and concluded that one factor is order—in *grego comerciante* 'Greek merchant', the noun is certainly *grego*; in *comerciante grego* it is *comerciante*. They also advanced the hypotheses—far more interesting from a semantic point of view—that the word classes for both noun and adjective, where tradition has packaged semantically diverse objects, are heterogeneous, and that our intuition regarding what is the nucleus of NPs containing a noun and an adjective is based on a combination of properties.

3.2 Gerunds

Portuguese has pairs of cognates such as *flutuante*/*flutuando*, *estudante*/*estudando*, *dormente*/*dormindo*, *reinante*/*reinando* etc. The words in these pairs derive from the same verbal root, but they have different histories (the words ending in *-nte* are the product of an erudite derivation; the words in *-ndo* had a popular derivation), and language seems to have classified for good the words ending in *-nte* as nouns or adjectives (for example, a pump working at the surface of water is called *bomba flutuante*; *dormente* 'sleeper' is a wooden crossbar used to support the rails of a railroad). The forms ending in *-ndo* maintain their link with the verb. According to traditional grammar, sentences such as *Vi o João estudando para a prova* 'I saw João studying for the test' or *João aprendeu inglês lendo romances* 'João learned English reading novels' illustrate the functions that Latin assigned to gerunds and participles, and these functions would have survived in Portuguese gerund. The reality of Portuguese, however, is more complex, and difficulties arise if we try to interpret *-ndo* words as ancient Latin participles when they follow *estar* (cf. Campos 1980).

3.3 Ser vs. Estar

Ser and *estar* occur quite often as copulae (*João é doente*; *João está doente* 'João is sick') or as location verbs. Explaining what distinguishes these verbs in each use is a long-standing problem for Portuguese (as well as other Romance languages). According to tradition, the difference would be the transient or permanent nature of the property denoted by the words following the copula: *a porta está aberta*/*a porta é de madeira* 'The door is open' vs. 'The door is made of wood', an explanation that should be extended to location sentences: *o helicóptero está no hangar*/*a delegacia é na esquina* 'The helicopter is in the hangar' vs. 'The police station is at the corner'. Once again, traditional explanations give an interesting but incomplete account of the facts. As for the usage of *ser* and *estar* in location sentences, Lemos (1987) noted that *ser* can take on a deontic reading: a housewife will say to the workers of a moving company: *A ca-*

deira de balanço é aqui 'The rocking chair is here' to indicate the place where she wants the rocking chair to be put. The choice between *ser* and *estar* may depend on the way one represents the place where the sentence is uttered. If a conversation is happening on a freeway along the coast, *Neste momento, o mar está a 12 quilômetros daqui* is more appropriate than *Neste momento, o mar é a 12 quilômetros daqui* 'Right now, the sea is 7 miles away' to indicate the distance of the sea.[19]

3.4 Passive *se*

Latin language formed passive sentences through the use of two devices which were applied to present and past tenses: the first employed specific verb endings (*boni scriptores laudantur*), the second a compound expression consisting of *esse* plus a past participle (*boni scriptores laudati sunt*). Portuguese grammarians also identified two different ways of obtaining passive sentences: one consists of a periphrasis based on *ser* (*os bons escritores são elogiados* 'good writers are praised'); the other appeals to the old reflexive pronoun *se*, converted into a passive operator (*elogiam-se os bons escritores*—same translation). Given two solutions in Latin and two solutions in modern Portuguese, only a short step was needed to suggest a homology between the two languages. The step was made when it was decided that Portuguese, similar to Latin, has both a synthetic and an analytical passive form. But the similarity of Portuguese and Latin here is no more than a wishful thinking—the two passive voices of Portuguese don't have equivalent meanings, as the Brazilian grammarian Said Ali remarked as early as 1908.[20]

Ali's analysis of the passive *se* is a remarkable example of semantic reflection applied to grammatical structures. His analysis ignored the debates of his time (mainly concerned with agreement, as *aluga-se casas* was already competing with *alugam-se casas*, both sentences meaning 'houses for rent') and opened a manner of explaining the role of *se* without falling into the inherited alternatives. When one does not look at the passive construction with *se* as an equivalent of Latin passive synthetic voice, it becomes easier to investigate what the different functions of *se* have in common.[21]

19 Schmitz (1977) highlights many interesting insights on this matter, contrasting Portuguese and English.

20 "*Aluga-se esta casa* and *esta casa é alugada* express two thoughts, differing in form and meaning [...] Let's post in front of a building an outdoor with the first of these phrases, let's post in front of another building the notice containing the caption "esta casa é alugada". The prospective tenants go without any doubt only to one of the houses, taking for granted that the other is already taken [...]" (Said Ali, ^{7}2008; original translation).

21 In tradition grammar, it has been more usual to look for differences than for similarities. Indeed, grammarians usually distinguish five meanings for the word *se*: passive, reflexive, reciprocal, indeterminate, and conditional conjunction.

Following this trend, the Brazilian linguist Carlos Franchi (1932–2001) used to suggest in his teaching that *se* occurs with transitive verbs that, instead of taking as a subject and object two different arguments, dispense with one of them because it is convenient to omit the agent (as in *Vendem-se casas*), because the same referent is both the agent and the *undergoer* (as in *João se vê no espelho* 'João sees himself in the mirror'), or because the grammatical subject is a plural that refers jointly to both (as in *Os dois amigos se cumprimentaram* 'The two friends greeted one another'). This hypothesis places *se* into the framework of the realization of verb arguments, an area where a certain instability is noticed in modern BP. Even if incorrect, Franchi's hypothesis is a reminder that the semantics of verb constructions, in spite of years of investigation, still needs attention.

4 A possible semantic agenda: new and old topics

One of the lessons to be learned by consulting the authors who did semantics *sans le savoir* is that their most interesting reflections were on aspects of language perceived as challenges. Like all languages, BP is rife with numerous theoretical and descriptive open issues which are exemplified below by considering two topics that have called the attention of semanticists in recent years: gerundism and bare singular nouns. In both cases, the point is to semantically explain (i.e., expliciting their meaning) two structures which are widely used in modern BP, in contrast to other Romance languages and (to some extent) in contrast to EP.

4.1 Gerundism

Gerundism is exemplified by phrases such as *Amanhã mesmo vamos estar mandando sua encomenda pelo correio* or *Para obter o desconto, o Sr. precisa estar apresentando este vale* 'Tomorrow, we'll be sending your order by mail'; 'To get the discount, you need to be presenting this voucher'. The main characteristic of gerundism is the fact that the infinitive *estar* 'to be' plus the gerund of a full verb occupy the place traditionally reserved for the infinitive of the last verb (*vamos estar mandando* instead of *vamos mandar* 'we'll send', *precisa estar apresentando* instead of *precisa apresentar* 'must present'). What nuance of meaning is expressed by this use of *estar*? It is well-known that the periphrasis formed with *estar* + gerund is old in BP (and is also one of the hallmarks of BP in contrast to EP, where the periphrasis *estar a* + infinitive prevails) and has always been used to indicate an ongoing action whereas simple forms of the verb indicate habit or repetition. Most scholars who have tried to explain the new structure have concluded that it is unnecessary, but this is not a semantic explanation. If, as postulated, gerundism appeared in the language as a significant new option, it is a good example of a new challenge to be tackled by semanticists. For recent accounts of

the phenomenon and the polemics surrounding it, see Possenti (2005) and Schmitz (2006).[22]

4.2 "Bare singular nouns"

Besides the controversy caused by recent innovations in language, complex semantic problems can be found in countless ways. One is the fruitful and critical coexistence of important linguistic theories, particularly those which have proved useful for other languages. This was precisely the story of the study of bare singular nouns in BP, which began to be investigated from a syntactic perspective for having an uncertain categorical status (DP, NP, or N); afterwards, these nouns were studied from a semantic perspective, and scholars discovered that they did not easily fit into the categories proposed by Carlson (1977) or Chierchia (1998) for their English analogues. BP bare singular nouns appear in sentences which do not accept verbatim translation in other languages, as exemplified in (8–13) below:

(8) Cachorro late (Literally: 'Dog barks', i.e., 'Dogs bark')
(9) *Chien aboie (but: Les chiens ça aboie)
(10) *Dog barks
(11) Dogs bark
(12) Japonês e mulher de cabelo curto guia mal (Literally: 'Japanese and short haired woman drive badly', i.e., 'Japanese men and short haired women drive badly')
(13) Português padeiro prefere mulata (Literally: 'Portuguese baker prefers mulatto woman', i.e., 'Portuguese bakers prefer mulatto women')

This difficulty of translation contradicts the claim that bare singular nouns in BP denote kinds and establishes a semantic puzzle that has been the subject of important studies.[23] This field of research has proved larger than was first suspected, as bare nouns are found not only in Portuguese but also in some Brazilian native languages such as Karitiana (Müller et al. 2006; Müller 2012). To what extent the syntactic representations and truth-conditions proposed for native languages apply to BP (and vice versa) is still an open question. The hypothesis that bare singular nouns have a generic interpretation must be carefully evaluated, because it turns a semantic puzzle into a

22 For a survey of the semantics (and pragmatics) of mood and modality, contrasting BP and EP, cf. Marques/Oliveira (2016).

23 Some of these studies can be found in Oliveira/Mezari (2012); Kabatek/Wall (2013); and Wall (2017). See, in particular, Oliveira (2010), Oliveira/Rothstein (2011), and Oliveira/de Swart (2015). On the semantics of singular bare nouns in BP, cf. Oliveira (2014). Ferreira/Correia (2016) present a contrastive semantics analysis of the DP in BP and in EP.

philosophical one—what is a kind? If "Japanese man and short haired woman" is a kind, it seems plausible that anything can be a kind. Perhaps bare nouns are only used in certain types of text; perhaps they express a kind of generalization that requires a particular manner of truth verification.[24]

4.3 And what about diachrony?

The semantic study of diachrony remains one of the major gaps of linguistic research on BP. This gap has been identified in various contexts, particularly in the *Projeto de História do Português Brasileiro* ('Project for the History of Brazilian Portuguese') that, based on written sources, attempts at giving the clearest possible account of changes occurred in BP from the colonial period to the present. Some of the papers already produced by this project deal with semantic issues via recent theories about language changes, such as Grammaticalization and Discursive Traditions—see, for example, Castilho da Costa (2011) and Longhin (2012). A comprehensive account of the material collected from the point of view of lexical and semantic changes which affected BP over five centuries is still missing, however. These changes deserve to be independently studied because they were responsible for several differences between BP and EP. A key step in this direction is being taken now through the collecting and digitizing of a vast amount of material.[25]

5 Conclusions

Semantics in Brazil has benefitted from a long process of imports that started in the 1960s. Semantic theories proposed in other places were assimilated with varying degrees of success by Brazilian researchers, in accordance with changes to the main guidelines in linguistic studies.

More than 50 years since this beginning, linguistics in Brazil faces not only a landscape of fragmentation and disputes but also large differences in the vitality of various orientations that have taken hold in Brazil. In this chapter, we have evaluated the vitality of theories not by the number of supporters but instead by the position they occupy in a scale that ranges from the contact with a new model to the emergence of broader analyses, through a stage of adequate assimilation and a critical interaction with foreign precursors. According to this scale, only some models are currently suc-

24 Bonomi (1975, 66–95) distinguishes between categorical generalization and non-categorical generalization. Many BP bare noun sentences seem to be of the later type.

25 On the diachronic semantics of BP, see the works collected in Ilari/Basso (2020), one the many works that results from the Projeto de História do Português Brasileiro.

cessful in producing original knowledge (mainly Formal, Cognitive, and Lexical Semantics).

For the reasons outlined above, we have returned to Brazilian scholars who practiced semantics *sans le savoir*. These scholars' contributions are encouraging, although much work is needed to recover old and scattered discussions that have little to do with semantic schools. We have also claimed that Portuguese, like any language, offers semanticists old and new challenges, in a repertoire that is apparently inexhaustible. Brazilian semanticists should look to this repertoire in the coming years, keeping in mind that while it is impossible to have good intuitions without a solid theoretical background, it can sometimes be useful to avoid building fences prematurely.

6 References

Ali, Manuel Said ([7]2008, [1]1908), *O pronome "se"*, in: *Dificuldades da Língua Portuguesa. Estudos e observações*, Rio de Janeiro, Academia Brasileira de Letras/Biblioteca Nacional, http://www.academia.org.br/abl/media/Dificuldades%20da%20Lingua%20Portuguesa%20-%20CAMS%20-%20PARA%20INTERNET.pdf, last accessed 18.02.2022.

Barreto, Terezinha Maria Mello (1999), *Conjunções: aspectos da sua constituição e funcionamento na história do português*, 3 vol., Doctoral thesis, Salvador, Universidade Federal da Bahia.

Basso, Renato Miguel (2007), *Telicidade e detelicização: semântica e pragmática do domínio tempo-aspectual*, Doctoral thesis, Campinas, UNICAMP.

Basso, Renato Miguel (2019a), *"Ao contrário", "do avesso" e "de trás para frente",* Cadernos de Estudos Linguísticos 61, 1–16, https://doi.org/10.20396/cel.v61i0.8654586.

Basso, Renato Miguel (2019b), *Sobre a semântica de "de volta": um exercício em contradirecionalidade*, Alfa, São Paulo 63, n. 2, 373–395.

Benveniste, Émile (1966), *Problèmes de linguistique générale*, Paris, Gallimard.

Benveniste, Émile (1970), *L'appareil formel de l'énonciation*, Langages 5/17, 12–18.

Bertucci, Roberlei Alves (2011), *Uma análise semântica para verbos aspectuais em português brasileiro*, Doctoral thesis, São Paulo, Universidade de São Paulo.

Boléo, Manuel de Paiva (1937), *O pretérito perfeito em português em confronto com as outras línguas românicas*, Coimbra, Imprensa da Universidade.

Bonomi, Andrea (1975), *Le vie del riferimento*, Milano, Bompiani.

Bonomi, Andrea/Usberti, Gabriele (1971), *Sintassi e semantica nella grammatica trasformazionale*, Milano, Mondadori [Brazilian translation: (1982), *Sintaxe e semântica na gramática transformacional*, São Paulo].

Borba, Francisco da Silva (2002), *Dicionário de Usos do Português do Brasil*, São Paulo, Ática.

Borba, Francisco da Silva, et al. (1990), *Dicionário de usos do português do Brasil*, São Paulo, Ática.

Borges Neto, José (1991), *Adjetivos: Predicados Extensionais e Predicados Intensionais*, Campinas, EDUNICAMP.

Bueno, Francisco da Silveira ([4]1965, [1]1947), *Tratado de Semântica Brasileira*, São Paulo, Saraiva.

Camacho, Roberto Gomes (2011), *As classes de palavras na perspectiva da Gramática Discursivo-Funcional. O papel da nominalização no continuum categorial*, São Paulo, EDUNESP.

Câmara Jr., Joaquim Mattoso (1975), *História e estrutura da língua portuguesa*, Rio de Janeiro, Padrão.

Campos, Odette de Sousa (1980), *O Gerúndio no Português*, Rio de Janeiro, Presença/INL-MEC.
Cançado, Márcia (2012), *Manual de Semântica. Noções básicas e exercício*, São Paulo, Contexto.
Cançado, Márcia/Godoy, Luisa/Amaral, Luana (2013), *Catálogo de verbos do português brasileiro, classificação verbal segundo a decomposição dos predicados*, vol. 1: *Verbos de mudança*, Belo Horizonte, EDUFMG.
Cançado, Márcia/Gonçalves, Anabela (2016), *Lexical Semantics: Verb Classes and Alternation*, in: W. Leo Wetzels et al. (edd.), *The Handbook of Portuguese Linguistics*, New York, Wiley-Blackwell, 374–391.
Carel, Marion (1992), *Vers une formalisation de la théorie de l'argumentation dans la langue*, Doctoral thesis, Paris, EHESS.
Carlson, Gregory Norman (1977), *Reference to kinds in English*, Doctoral thesis, Amherst (MA), University of Massachusetts.
Castilho da Costa, Alessandra (2011), *Tradições discursivas em jornais paulistas de 1854 a 1901: Gêneros entre a história da língua e a história dos textos*, Munich, GRIN Verlag.
Castilho, Ataliba Teixeira de (1967), *Introdução ao estudo do aspecto verbal na língua portuguesa,* Marília, Coleção de Teses.
Castilho, Ataliba Teixeira de (2007), *An approach to language as a complex system. New issues in Historical Linguistics*, Signos Linguísticos 6, 83–120.
Castilho, Ataliba Teixeira de (2009), *An approach to language as a complex system*, in: Ataliba Teixeira de Castilho (ed.), *História do Português Paulista*, vol 1, Campinas, IEL, 119–136.
Castilho, Ataliba Teixeira de (2010), *Nova gramática do português brasileiro*, São Paulo, Contexto.
Chierchia, Gennaro (1998), *Reference to Kinds across Languages*, Natural Language Semantics 6, 339–405.
Chierchia, Gennaro/McConnell-Ginet, Sally (1990), *Meaning and Grammar: An Introduction to Semantics*, Boston (MA), MIT Press.
Cunha Lima, Maria Luiza (2013), *Semântica e Psicolinguística Experimental*, in: Celso Ferrarezi Jr./ Renato M. Basso (edd.), *Semântica, Semânticas*, São Paulo, Contexto, 121–134.
Dascal, Marcelo (1982a), *Relevância Conversacional*, in: *Fundamentos Metodológicos da Linguística*, vol. 4: *Pragmática*, Campinas, Edição do autor, 105–133. [Originally published in: (1977), *Conversational Relevance*, Journal of Pragmatics 1, 309–328; republished in: (2006a), *Filosofia e Ciência, Interpretação e Compreensão*, São Leopoldo, Unisinos, 54–74].
Dascal, Marcelo (1982b), *Comecemos a acabar de começar? Prolegômenos para uma análise semântica de algumas perífrases verbais indicadoras de fase, em português*, Cadernos de Estudos Linguísticos 3, 123–186.
Dascal, Marcelo (ed.) (1982c), *Fundamentos metodológicos da linguística*, vol. 3: *Semântica*, Campinas, Edição do autor.
Dascal, Marcelo (2006a), *Filosofia e Ciência, Interpretação e Compreensão*, São Leopoldo, Unisinos.
Dascal, Marcelo (2006b), *Interpretação e Compreensão*, São Leopoldo, RS, Editora Unisinos.
Davidson, Donald (1967a), *Truth and Meaning*, Synthese 17, 304–323.
Davidson, Donald (1967b), *The Logical Form of Action Sentences*, in: Nicholas Rescher (ed.), *The Logic of Decision and Action*, Pittsburgh, University of Pittsburgh Press.
Dowty, David (1979), *Word meaning and Montague Grammar. The semantics of Verbs and Times in Generative Semantics and in Montague's PTQ*, Dordrecht, Reidel.
Dowty, David (1991), *Thematic Proto-Roles and Argument Selection*, Language 67/3, 547–619.
Ducrot, Oswald (1972), *Dire et ne pas dire. Principes de sémantique linguistique*, Paris, Hermann [Brazilian translation: (1977), *Dizer e não dizer. Princípios de semântica lingüística*, São Paulo, Cultrix].
Ducrot, Oswald (1980), *Le dire et le dit*, Paris, Minuit.

Ducrot, Oswald/Vogt, Carlos Alberto (1979), *De "magis" à "mais", une hypothèse sémantique*, Revue de Linguistique Romane 43, 317–341.
Fauconnier, Gilles (1994), *Mental Spaces*, Cambridge, MIT Press.
Ferrarezi Jr., Celso (2013), *Semântica Cultural*, in: Celso Ferrarezi Jr./Renato M. Basso (edd.), *Semântica, Semânticas*, São Paulo, Contexto, 71–83.
Ferrarezi Jr., Celso/Basso, Renato M. (2013), *Semântica, Semânticas*, São Paulo, Contexto.
Ferreira, Marcelo Barra (2019), *Curso de semântica formal*, Berlin, Language Science Press.
Ferreira, Marcelo Barra (2004), *Imperfectives and Plurality*, in: *Proceedings of SALT 14*, Ithaca (NY), CLC Publications,1–18.
Ferreira, Marcelo Barra/Correia, Clara Nunes (2016), *The Semantics of DPs*, in: W. Leo Wetzels et al. (edd.), *The Handbook of Portuguese Linguistics*, New York, Wiley-Blackwell, 356–373.
Figueira, Rosa Attié (1974), *Verbos Introdutores de Pressupostos*, Doctoral thesis, Campinas, UNICAMP.
Fillmore, Charles (1966), *The case for case*, in: Emmon W. Bach/Robert T. Harms (edd.), *Universals in Linguistic Theory*, London, Holt, Rinehart and Winston, 1–88.
Fillmore, Charles (1977), *The case for case reopened*, in: Peter Cole/Jerry M. Sadok (edd.), *Syntax and Semantics*, vol. 8, New York, Academic Press, 59–81.
Fillmore, Charles/Kay, Paul/O'Connor, Catherine (1988), *Regularity and Idiomaticity in Grammatical Constructions: The Case of "let alone"*, Language 64, 501–38.
Fiorin, José Luiz (2002), *Adjetivos temporais e espaciais*, in: Maria Bernadete M. Abaurre/Angela C. S. Rodrigues (edd.), *Gramática do Português Falado*, vol. 8: *Novos estudos descritivos*, Campinas, EDUNICAMP, 59–81.
Foltran, Maria José (2011), *Os advérbios de modo em "-mente" e a estrutura do evento*, in: Roberta Pires de Oliveira/Carlos Mioto (edd.), *Percursos em Teoria da Gramática*, Florianópolis, EDUFSC, 165–182.
Franchi, Carlos (2012/1982), *Relatório de atividades referente ao afastamento de 15/12/80 a 30/9/81 para desenvolver atividades de estudo e pesquisa na U.C. Berkeley*, Revista da ABRALIN 2012/2, 229–337.
Geerarts, Dirk (1997), *Diachronic Prototype Semantics, a contribution to historical lexicology,* Oxford, Clarendon Press.
Geerarts, Dirk (2010), *Theories of Lexical Semantics*, Oxford, Oxford U.P.
Geraldi, João Wanderley (1978), *Se a semântica fosse também pragmática. Para uma análise semântica dos enunciados condicionais*, Master's thesis, Campinas, UNICAMP.
Goldberg, Adele (1995), *Constructions: A Construction Grammar Approach to Argument Structure*, Chicago, UCP.
Gomes, Ana Paula Quadros (2009), *O efeito grau máximo sobre os domínios*, Doctoral thesis, São Paulo, Universidade de São Paulo.
Gomes, Ana Quadros/Mendes, Luciana Sanchez (2018), *Para conhecer Semântica*, São Paulo, Contexto.
Gonçalves, Cláudio C. e C. (2007), *Imperfectivity and Verbal Morphology*, Doctoral thesis, Florianópolis, Universidade Federal de Santa Catarina.
Gonçalves, Rodrigo Tadeu/Pagani, Luiz Arthur (2004), *O efeito labirinto além da sintaxe. Eliminando a ambiguidade,* Revista Letras 63, 177–195.
Gonçalves, Sebastião Carlos Leite/Lima-Hernandez, Maria Célia/Casseb-Galvão, Vânia Cristina (edd.) (2007), *Introdução à gramaticalização: princípios teóricos e aplicação*, São Paulo, Parábola.
Guimarães, Eduardo R. J. (1987), *Texto e argumentação: um estudo de conjunções no português*, Campinas, Pontes.
Guimarães, Eduardo R. J. (2004), *História da Semântica: Sujeito, sentido e Gramática no Brasil*, Campinas, Pontes.

Guimarães, Márcio Renato (2002), *Alguns problemas na interpretação da progressividade no português do Brasil*, Revista Letras 58, 185–209.
Heim, Irene/Kratzer, Angelika (1998), *Semantics in generative Grammar*, New York, Wiley-Blackwell.
Holanda, Aurélio Buarque de, et al. (1938), *Pequeno Dicionário Brasileiro da Língua Portuguesa*, Editora do Autor.
Ilari, Rodolfo (1998a), *Pela primeira vez e suas complicações sintático-semânticas*, DELTA 14/2, 133–144.
Ilari, Rodolfo (1998b), *Notas para uma semântica do passado composto em português*, in: *Atas do Congresso Internacional organizado por motivo dos Vinte anos do Português no ensino superior*, Budapest, Eötvós Loránd University, 224–247.
Ilari, Rodolfo (2001), *Introdução à semântica: brincando com a gramática*, São Paulo, Contexto.
Ilari, Rodolfo (2002), *Introdução ao estudo do Léxico: brincando com as palavras*, São Paulo, Contexto.
Ilari, Rodolfo/Geraldi, João Wanderley (1985), *Semântica*, São Paulo, Ática.
Ilari, Rodolfo/Basso, Renato Miguel (2014), *O verbo*, in: Rodolfo Ilari (ed.), *Palavras de Classe Aberta,* São Paulo, Contexto, 65–241.
Ilari, Rodolfo/Basso, Renato Miguel (2020), *História do Português Brasileiro, vol. 8: História Semântica do Português Brasileiro*, São Paulo, Contexto.
Ilari, Rodolfo/Oliveira, Maria Fátima/Basso, Renato Miguel (2016), *Tense and Aspect: a Survey*, in: W. Leo Wetzels et al. (edd.), *The Handbook of Portuguese Linguistics*, New York, Wiley-Blackwell, 392–407.
Ilari, Rodolfo, et al. (1996), *Considerações sobre a posição dos advérbios*, in: Ataliba Teixeira de Castilho (ed.), *Gramática do português falado*, vol. 1: *A Ordem*, Campinas, Edunicamp, 63–141.
Kabatek, Johannes/Wall, Albert (2013), *New Perspectives on Bare Noun Phrases in Romance and Beyond*, Amsterdam, Benjamins.
Katz, Jerrold J./Fodor, Jerry A. (1963), *Structure of a semantic theory*, Language 39/2, 170–210 [Brazilian translation: *Estrutura de uma teoria semântica*, in: Lúcia Maria Pinheiro Lobato (ed.) (1977), *A semântica na linguística moderna: o léxico*, Rio de Janeiro, Francisco Alves, 77–129].
Koch, Ingedore (2002), *Desvendando os segredos do texto*, São Paulo, Cortez.
Lakoff, George (1982), *Categories and cognitive models*, Berkeley, California, Berkeley Cognitive Science Program.
Lemos, Claudia T. G. (1987), *"Ser" and "estar" in Brazilian Portuguese with particular reference to child language acquisition*, Tübingen, Narr.
Lenz, Paula (2013), *Semântica cognitiva*, in: Celso Ferrarezi Jr./Renato M. Basso (edd.), *Semântica, Semânticas*, São Paulo, Contexto, 31–55.
Lobato, Lúcia Maria Pinheiro (1970), *L'auxiliarité en langue portugaise*, Doctoral thesis, Paris, Université de Paris.
Lobato, Lúcia Maria Pinheiro (ed.) (1977), *A semântica na linguística moderna: o léxico*, Rio de Janeiro, Francisco Alves.
Lobato, Lúcia Maria Pinheiro (1993), *Adjetivos: uma tipologia e interpretação semântica*, Boletim da ANPOLL 14, 9–25.
Longhin, Sanderléia Roberta (2003), *A gramaticalização da perífrase conjuncional 'só que',* Doctoral thesis, Campinas, UNICAMP.
Longhin, Sanderléia Roberta (2012), *Junção, tradição discursiva e mudança linguística*, in: Manoel Mourivaldo Santiago-Almeida (ed.), *História do Português Paulista*, vol. 3, Campinas, IEL, 367–380.
Lopes dos Santos, Luiz Henrique (2008), *O olho e o microscópio: a gênese e os fundamentos da lógica segundo Frege*, Rio de Janeiro, Nau Editora.

Lyons, John (ed.) (1976), *Novos Horizontes em Linguística*, São Paulo, Cultrix [Original: ([1]1970), *New horizons in linguistics*, Harmondsworth, Penguin].
Lyons, John (1977), *Semantics*, 2 vol., Cambridge, CUP
Marchesan, Ani Carla (2011), *A sintaxe das relativas livres no Português Brasileiro*, in: Roberta Pires de Oliveira/Carlos Mioto (edd.), *Percursos em Teoria da Gramática*, Florianópolis, EDUFSC, 73–94.
Martins, Ronaldo Teixeira (2004), *A língua nova do imperador*, Doctoral thesis, Campinas, UNICAMP.
Marques, Rui/Oliveira, Roberta Pires de Oliveira (2016), *Mood and modality*, in: W. Leo Wetzels et al. (edd.), *The Handbook of Portuguese Linguistics*, New York, Wiley-Blackwell, 408–424.
Menuzzi, Sérgio de Moura (2018), *Sobre a pressuposição das clivadas*, Revista da ANPOLL 46, 200–221.
Menuzzi, Sérgio de Moura (2012), *Algumas Observações sobre Foco, Contraste e Exaustividade*, Revista Letras 86, 95–121.
Menuzzi, Sérgio de Moura/Rodrigues, Gabriel Roisenberg (2010), *Tópicos Contrastivos e Contraste Temático: Um Estudo do Papel Discursivo da "Articulação Informacional",* Cadernos de Estudos Linguísticos (UNICAMP) 52, 233–253.
Molsing, Karina (2010), *The present-perfect: an exercise in the study of events, plurality and aspect*, Doctoral thesis, Porto Alegre, PUC-RGS.
Müller, Ana Lúcia de Paula (2002), *The Semantics of Generic Quantification in Brazilian Portuguese*, PROBUS 14, 279–298.
Müller, Ana Lúcia de Paula (2004), *Tipos de Genericidade como Tipos de Modalidade*, Revista Letras 60, 341–362.
Müller, Ana Lúcia de Paula (2012), *Distributividade: o caso dos numerais reduplicados em Karitiana*, Cadernos de Estudos Linguísticos 54, 225–243.
Müller, Ana Lúcia de Paula (ed.) (2003), *Semântica formal*, São Paulo, Contexto.
Müller, Ana Lúcia de Paula, et al. (2006), *Número e a Distinção Contável-Massivo em Karitiana*, Revista da ABRALIN 5, 185–213.
Negri, Ligia (2006), *Zona de fronteira: a delimitação entre a semântica e a pragmática sob a lente das expressões de polaridade negative*, Doctoral thesis, Curitiba, UFPR.
Negri, Ligia, et al. (edd.) (2004), *Sentido e Significado: em torno da obra de Rodolfo Ilari*, São Paulo, Contexto.
Oliveira, Roberta Pires de (2001), *Semântica formal, uma breve introdução*, Campinas, Mercado de Letras.
Oliveira, Roberta Pires de (2010), *Bare Noun Phrases in focus*, Journal of Portuguese Linguistics 9, 3–8.
Oliveira, Roberta Pires de (2014), *Dobras e Redobras: do singular nu no português brasileiro: costurando a semântica entre as línguas*, Porto Alegre, EDIPUCRS.
Oliveira, Roberta Pires de/Rothstein, Susan (2011), *Bare singular noun phrases are mass in Brazilian Portuguese*, Lingua 121, 2153–2175.
Oliveira, Roberta Pires de/de Swart, Henriette (2015), *Brazilian Portuguese Noun Phrases: an optimality theoretic perspective*, Journal of Portuguese Linguistics 13, 63–94.
Oliveira, Roberta Pires de/Mezari, Meiry Peruchi (edd.) (2012), *Nominais nus, um olhar através das línguas*, Campinas, Mercado de Letras.
Oliveira, Roberta Pires de/Mioto, Carlos (edd.) (2011), *Percursos em Teoria da Gramática*, Florianópolis, EDUFSC.
Oliveira, Roberta Pires de/Basso, Renato Miguel (2014), *Arquitetura da conversação: teorias das implicaturas*, São Paulo, Parábola.
Oliveira, Roberta Pires de/Emmel, Ina/Quarezemin, Sandra (2020), *Brazilian Portuguese, Syntax and Semantics: 20 years of Núcleo de Estudos Gramaticais*, Amsterdam, Benjamins.

Pagani, Luiz Arthur (1998), *Gramática categorial e processamento linguístico*, Revista Letras 50, 171–183.
Poggio, Rosauta (2002), *Processos de gramaticalização de preposições do latim ao português; uma abordagem funcionalista*, Salvador, EDUFBA.
Pontes, Eunice (1971), *Verbos auxiliares em português*, Petrópolis, Vozes.
Possenti, Sírio (2005), *Defendendo o gerúndio*, Discutindo Língua Portuguesa 1, 8–11.
Reichenbach, Hans (1947), *Elements of Symbolic Logic*, London, MacMillan.
Rosch, Eleanor (1973), *On the Internal Structure of Perceptual and Semantic Categories*, in: T. Moore (ed.), *Cognitive Development and the Acquisition of Language*, New York, Academic Press.
Schmitz, John Robert (1977), *Análise contrastiva de erros: "Estar Com"; "Ter" em Português e "Be", "Have" em inglês*, Revista Brasileira de Linguística 4/1, 38–48.
Schmitz, John Robert (2006), *Sobre o gerúndio e "gerundismo": uma análise de um assunto emotivo e polêmico*, Confluência 31, 87–110.
Silva, Augusto Soares da (2006), *O mundo dos sentidos em português, polissemia, semântica e cognição*, Coimbra, Livraria Almedina.
Souza, Luisandro Mendes de (2010), *Comparativas Quantificacionais no português brasileiro: semântica e sintaxe,* Doctoral thesis, Florianópolis, Universidade Federal de Santa Catarina.
Souza, Luisandro Mendes de (2011), *Considerações sobre a sintaxe das sentenças comparativas*, in: Roberta Pires de Oliveira/Carlos Mioto (edd.), *Percursos em Teoria da Gramática*, Florianópolis, EDUFSC, 115–132.
Taffarello, Cristina (1979), *"Mesmo": uma proposta de análise dentro da teoria da semântica argumentativa*, Master's thesis, Campinas, UNICAMP.
Taylor, John R. (1989), *Linguistic Categorization*, Oxford, Clarendon.
Teixeira, Mariana T./Menuzzi, Sérgio de Moura (2015), *Diverse "Exhaustiveness Effects" in Clefts*, Alfa: Revista de Linguística 59, 59–87.
Travaglia, Luiz Carlos (1981), *O aspecto verbal em português*, Uberlândia, Editora da UFU.
Vendler, Zeno (1967), *Linguistics and Philosophy*, Ithaca, Cornell U.P.
Viana, Aniceto dos Reis Gonçalves (1901), *Le Portugais*, Leipzig, Teubner.
Vogt, Carlos Alberto (1977), *O intervalo semântico*, São Paulo, Ática.
Vogt, Carlos Alberto (1980), *Linguagem, Pragmática e Ideologia*, São Paulo, Hucitec.
Votre, Sebastião Josué (1994), *Projeto integrado, Iconicidade, cognição e gramaticalização*, CNPq.
Wachowicz, Tereza Cristina (2003), *A sintaxe e a semântica do progressivo em português brasileiro*, Doctoral thesis, São Paulo, Universidade de São Paulo.
Wachowicz, Tereza Cristina (2013), *Semântica Lexical*, in: Celso Ferrarezi Jr./Renato M. Basso (edd.), *Semântica, Semânticas*, São Paulo, Contexto, 153–170.
Wall, Albert (2017), *Bare nominals in Brazilian Portuguese: an integral approach*, Amsterdam/Philadelphia, Benjamins.
Weinreich, Uriel (1966), *Explorations in Semantic Theory*, in: Thomas A. Sebeok (ed.), *Current Trends in Linguistics*, The Hague, Mouton, 395–477 [*Pesquisas em teoria semantica*, in: Lúcia Maria Pinheiro Lobato (ed.) (1977), *A semântica na linguística moderna: o léxico*, Rio de Janeiro, Francisco Alves, 165–273].
Wetzels, W. Leo/Costa, João/Menuzzi, Sergio (edd.) (2016), *The Handbook of Portuguese Linguistics*, New York, Wiley-Blackwell.
Wilson, Deirdre (1975a), *Presupposition and non-truth-conditional semantics*, London, Academic Press.
Wilson, Deirdre (1975b), *Presupposition and the delimitation of semantics*, Cambridge, CUP
Zamboni, Lilian M.S. (1978), *Análise semantica da palavra "ainda"*, Master's thesis, Campinas, UNICAMP.

Alessandra Castilho da Costa

17 Spoken *vs.* written language

Abstract: This chapter presents an overview of the oral and written production of texts throughout the course of Brazilian history, providing illustrative historical information. The central focus is on the identification of discourse traditions that arose as a response to the communicative needs of various cultural groups and institutions that constituted the Brazilian society in the past, and that arise in the present, as well as how these reflect different linguistic varieties. Such an approach does not rely on the evolution of the national languages as a framework, but rather on communicative situations and texts associated with it. In section 1 we will present the main theoretical concepts in this area of study (discourse genres, discourse tradition, orality and literacy, among others). The remaining sections comprise examples of texts and discourse traditions that play a central role in communication in Brazil, both in the past and now: at the beginning of the country's colonization (2), during the colonial and imperial period (3), in the 19th century (4), and from the 20th century to the present day (5).

Keywords: spoken texts, written texts, discourse tradition, orality, literacy, immediacy, distance

1 Discourse traditions and the orality-literacy continuum

A text, whether transferred phonically or graphically, not only activates the language-based rules of a linguistic system, but also the rules governing the organization of the text (Coseriu [3]1994, 51). Directly aligned with the theory proposed by Eugenio Coseriu, Koch (1997) develops the concept of *discourse traditions*. This term covers all forms of traditionality in texts, both written and oral: universes of discourse (constellations of broader discourse traditions, or DTs, such as literature, science, and religion), styles, conversational forms, and speech acts, among others. An initial distinction established by Koch involves the two previously mentioned speech traditions: the first of these relative to linguistic norms (language traditions), and the second relative to the discourse norms (text or discourse traditions). The fundamental difference between the two types of traditions resides in the respective group of 'transporters' (Germ. *Trägergruppen*). On the one hand, the rules of a *language* are transported by linguistic communities, and organize purely linguistic facts. On the other hand, rules of *discourse* are transported by cultural groups, such as professional groups, literary trends, political movements, etc.

To establish a systematic distinction between speech and writing, Koch/Oesterreicher ([2]2011) developed a theoretical model in which language manifests itself through

https://doi.org/10.1515/9783110405958-018

discourse traditions and linguistic varieties from a global concept of orality and literacy, that is, of *the language of immediacy* and *communicative distance*, respectively (following the terminology of the authors). With this model, all discourse traditions possess both a medial profile and a conceptual one. For example, the *personal letter* is graphically delivered, while *a pronouncement from the President of the Republic* is phonically delivered. As for the conception, a personal letter presents affinities with the language of immediacy and is correlated, therefore, to vernacular usage. The pronouncement of the President, due to the very formal and public nature of the communication, uses a linguistic variety that is diatopically neutral, as well as diastratic and diaphasic varieties which are considered prestigious. At the two poles of the continuum are the discourse traditions most prototypical of the language of immediacy, and of communicative distance. From a perspective that covers both *medium* and *conception*, different discourse traditions can be located between these two extremes (Koch/Oesterreicher [2]2011, 10–13; see also Kabatek 2006; 2018).

In this context of relations between discourse traditions and the orality and literacy continuum, the concept of *linguistic elaboration* (in German, *Ausbau*), proposed by Koch/Oesterreicher ([2]2011, 136), drawing on the ideas of Heinz Kloss ([2]1978), addresses the fact that linguistic varieties may advance towards new discourse traditions, particularly towards those of communicative distance. On the one hand, a linguistic variety that is as yet unwritten should be able to be employed in all the discourse traditions in use in the corresponding culture (extensive elaboration). However, the linguistic variety in question must develop linguistic structures that satisfy the requirements of communicative distance (intensive elaboration).

In what follows, the orality/literacy-continuum will be the background for the description of the course of the history of communication in Brazil in order to show how certain discourse traditions performed a notable role in communicative processes in different historical periods.

2 Oral and written texts in the early colonial period of Brazil

According to Faraco (2016, 120–123), from the 3rd decade of the 16th century, Portugal began its occupation of Brazilian territory with the goal of extracting resources and growing sugar cane, as well as converting the "pagan" population to Roman Catholicism, this through a Jesuit-led campaign of religious conversions. In order to catechize the Indigenous populations, the Jesuits removed people from their villages and installed them in settlements, where they received spiritual instruction from the Company of Jesus (↗2 The social history of Brazilian Portuguese).

Some discourse traditions performed a particularly significant role in the communicative processes of this period. Colonial historians claim that among those that cir-

culated orally within Indigenous tribes were fables, legends (which served as a type of theogony), myths (see Steinen 1894) and a unique poetic style in which the ends of the motes (refrains) of its chants were repeated (Cascudo 2006, 83–148):

(1) Yáputi ne maquyra, 'Tie your hammock,
Tamaquaré. Tamaquaré.
Cha quire putare uana, I want to sleep now,
Tamaquaré. Tamaquaré.
Ure uana coena, Dawn is coming,
Tamaquaré Tamaquaré.'
(source: Cascudo 2006, 148).

Moreover, according to the folklorist Câmara Cascudo, other discourse traditions were the *moacaretá* and the *poranduba*. Indigenous chiefs were accustomed to gathering around the fire to discuss tribal life. During the *moacaretá* (in the Tupi-Guarani language), a type of council of leaders, the young gained knowledge of warrior traditions and legends through instruction by the elders. The *poranduba* or *marunduba* is a narrative retelling of the daily life of an Indigenous person over the period of a day, or the carrying out of a particular task over many days; in other words, a type of orally expressed odyssey or journey of an individual, a tradition which continues in the interior regions of the country to this day (Cascudo 2006, 84; see also Rodrigues 1890; Stradelli 1929). The *moacaretá* and *poranduba* are also catalogued in documents by Hans Staden, André Thevét, Jean de Léry, Father Ivo d'Evreux, among others.

Nevertheless, the production of Indigenous oral literary traditions is reduced to the geographic areas in which particular tribes established themselves, while later Afro-Brazilian production spread rapidly through the *mestiços* (mixed race) and second-generation Brazilians, who were referred to as *caboclos* and *mulatos*, forming a means of transmitting the (hi)stories of their parents (Cascudo 2006, 34).

Many colonists, missionaries, Tupi speakers and the *mestiça* population were bilingual Portuguese-Tupi, or were 'general language' speakers. Lucchesi (2009, 43) describes how the innumerable languages found by the first Portuguese settlers in Brazil, although distinct, were closely related to one another, especially the languages from the Tupi group. For this reason, to communicate with the native population the Portuguese used "general languages", that is, inter-ethnic vehicular languages (↗2 The social history of Brazilian Portuguese) used by populations, especially the language copied from the Tupinambá, which served as a type of lingua franca, referred to as *the general language of the Brazilian coast*.

With respect to this, Silva (2004, 127) considers the term *lingua geral da costa*– 'general coastal language' to be imprecise, since at least two general languages were used, in the South and the North of the country, respectively; the general Paulista language based on Tupiniquim and Guarani, used by the *bandeirantes* (pioneers) in settling the interior of São Paulo and the Midwest, and the general Amazonian language,

based on Tupinambá, which was developed initially in Maranhão and Pará State, serving the colonization of the Amazon, from which come the *Nheengatu* of the Içana Valley (bordering Venezuela, ↗2 The social history of Brazilian Portuguese).

Despite any such imprecision as to the scope of the term, some discourse traditions register Indigenous and general languages from the early period of colonization. This mainly involves three text traditions: historiographic prose, spiritual catechism literature, and metalinguistic literature. Within historiographic prose we find letters, memoires and historical accounts by travelers, naturalists, and adventurers (not only Portuguese) who came to Brazil in the first decades of colonization, such as *Warhaftige Historia und Beschreibung eyner Landtschafft der Wilden Nacketen, Grimmigen Menschfresser-Leuthen in der Newenwelt America gelegen* 'True Story and Description of a Country of Wild, Naked, Grim, Man-Eating People in the New World America' (1557) by Hans Staden, *Les Singularités de la France Antarctique* 'Peculiarities of French Antarctica' (1557) by André Thevet, and *Histoire d'un voyage faict en la terre du Brésil* 'Travels through the land of Brazil' (1578) by Jean de Léry. This last work, for example, includes a dialogue with an Indigenous man, in both Tupi and French. In the area of metalinguistic literature, noteworthy works include grammars such as *Arte de gramatica da lingua mais usada na costa do Brasil* 'The art of grammar of the language most used on the coast of Brazil' (1595) by José de Anchieta, *Arte de grammatica da língua brasílica* 'The art of grammar of the Brasilica language' (1687) by Luís Figueira, *Língua geral dos índios das américas: Pernanbúque, Parâ, Marainhaõ, Bahia, Rio de Janeiro, S. Paullo e Minas Gera[is]* 'General language of the Indians of the Américas: Pernambuco, Pará, Maranhão, Bahia, Rio de Janeiro, São Paulo and Minas Gerais' (without date, probably from the 18th century) by an anonymous author, *Diccionario Portuguez-Brasiliano e Brasiliano-Portuguez* 'Dictionary of Portuguese-Brazilian and Brazilian-Portuguese' (1795) by an anonymous author, *Vocabulario na Lingua Brasilica* 'Vocabulary in the Brasilica Language', by an anonymous author from the 17th century, *Diccionario da Língua Geral do Brasil que se falla em todas as villas, lugares e aldeãs deste vastíssimo Estado* 'Dictionary of the General Language of Brazil spoken in all the villages, places and communities from the vast State' (1771), unknown author, *Glossários de diversas lingoas e dialectos, que fallao os índios no imperio do Brazil* 'Glossary of the diverse languages and dialects that the Indians of the Empire speak' (1867) by Karl Friedrich Philipp von Martius, among others. Meanwhile, the spiritual literature of catechism was also included in various records: *Dialogo da doutrina cristã pela língua brasílica* 'Dialogue of the Christian doctrine through the Brasilica language', supposedly written by Padre Luis Maria Bucherelli between 1718–1749, *Compêndio da doutrina cristã que se manda ensinar como preceito no anno de 1740* 'Compendium of the Christian doctrine that is taught as a precept in the year 1740', by Marcos Antonio Arnolfini, *Doutrina e perguntas dos mistérios principais de Nossa Santa Fé na língua brasílica* 'Doctrine and questions of the main mysteries of Our Holy Faith in the Brasilica language' by an unknown author and undated, *Catecismo na língua brasílica* 'Catechism in the Brasilica language' (1618) by Antônio de Araújo, *Catecismo Brasilico da doutrina christãa, com o cerimonial dos Sa-*

cramentos, & mais actos Parochiaes—'Brasilico Catechism of the Christian doctrine, with the Holy Sacraments and other parochial acts' (1686), also by Antônio de Araújo, *Compêndio da Doutrina Christam na Língua Portuguesa, e Brasilica*—'Compendium of the Christian doctrine in the Portuguese and Brasilica language' (1687) by João Felipe Bettendorff, among others.

From the point of view of the conceptual continuum, the Portuguese language functioned as a language of distance, used in the discourse traditions of administration and law by the administrative authorities, but also as a language of communicative immediacy in daily interactions between colonists. For this reason, in facing the extant linguistic diversity of the colony, it was necessary for the colonists and missionaries of the Company of Jesus to learn the general languages (Faraco 2016, 126). Efforts towards religious conversion required the transposition of these languages to religious discourse traditions, influenced by the language of distance. Due to the catechism being conducted in local languages, the Jesuit missionaries produced grammar books and bilingual vocabularies, both of the general languages and of the various other Indigenous languages. Thus, in the sphere of the religious DT, a corpus of spiritual literature emerged in various genres: catechisms, scripts for confessions, sermons, manuals, homilies, hagiologies, Christian psalms and religious songs, translations of the gospels and biblical epistles, papal bulls, missals, saints' biographies, *autos* (religious dramas) and religious poems (Barros 1995; see also Daher 1998, 33–34), among others, in which these general languages are registered.

In his catechizing mission, Father José de Anchieta, for example, produced at least 12 *autos* in Portuguese, Spanish and Tupi, in which Indigenous and European characters and themes were united. The following are some examples: *Auto da Pregação Universal* 'Act of Universal Preaching' (performed in 1561 and written in Portuguese, Tupi and Spanish), *Auto de São Lourenço* 'Act of São Lourenço' (performed in 1587, in Tupi), *Na Vila de Vitória ou Auto de São Maurício* 'In the village of Vitória or Act of São Maurício' (performed in 1595, in Portuguese and Spanish), *Auto de Ursula ou Quando no Espírito Santo se recebeu uma relíquia das onze mil virgens* 'Act of Ursula or When in Holy Spirit a relic of one thousand and one virgins was received' (performed in 1595, in Portuguese), *Auto da Crisma ou Recebimento do administrador apostólico P. Bartolomeu Simões Pereira* 'Act of Chrism or Reception of the Apostolic Administrator Father Bartolomeu Simões Pereira' (performed in 1591, in Portuguese and Tupi), *Auto de São Sebastião* 'Act of São Sebastião' (performed in 1584, in Tupi), *Diálogo do P. Pero Dias Mártir* 'Dialogue of Father Pero Dias Martyr' (performed between 1584 and 1585, in Spanish), *Na Aldeia de Guarapari* 'In the village of Guarapari' (performed in 1585, in Tupi), *Recebimento que fizeram os índios de Guarapari ao Padre Provincial Marçal Beliarte* 'Reception of Provincial Father Marçal Beliarte by the Indians of Guarapari' (performed in 1589, in Portuguese and Tupi), *Dia da assunção, quando levaram sua imagem a Reritiba* 'Assumption Day, when they took her image to Reritiba' (performed toward the end of 1591 and the beginning of 1592, in Tupi), *Recebimento do P. Marcos da Costa* 'The reception of Father Marcos da Costa' (1596, in

Portuguese and Tupi), and *Na Visitação de Santa Isabel* 'In the visitation of Saint Isabel' (1597, in Spanish).

In the socio-interactional framework of the 16th century, work to develop the Tupi language linguistically was carried out through the channel of the discourse traditions associated with Roman Catholicism. More precisely, linguistic resources were developed so that Tupi could also be used in written religious texts of communicative distance. This process of linguistic elaboration, which Faraco (2016, 134) characterizes as "the adjustment of the Tupi language to the catechist discourse (of merging the European religious memory to the Indigenous linguistic-cultural universe)", was achieved through the creation of neologisms, phraseological decals, loan words from Portuguese, among other strategies for the creation of the linguistic means necessary to carry out communicative tasks that the catechistic discourse demanded (↗1 Brazilian Portuguese linguistics: an overview).

In this context of Jesuit spiritual instruction, Barros (2003, 131) considers the most significant tradition to be the catechism, which has as its main organizing feature the adjacency pair question (by the master) and answer (by the disciple) where learning is done by memorizing what is heard and by repeating *verbatim* what is said. An example of this discourse tradition is the *Catecismo na língua brasílica* 'Catechism in the Brasilica language', organized by Father Antônio Araújo from the texts of his predecessors, such as Anchieta and Pero Correia, which saw its first edition printed in 1618. Nevertheless, in the 18th century, the Tupi daily language had become so different from the crystallized one in grammars and catechism texts that the Jesuit João Daniel (1722–1776), in *Tesouro descoberto no máximo rio Amazonas* (between 1757 and 1776) 'Treasure discovered in the largest river of the Amazon', recounts how the Indigenous population no longer understood the catechism. Likewise, those who studied the grammars and vocabularies no longer managed to communicate with Tupi speakers. For this reason, according to Faraco (2016, 134–135), in the 18th century, an anonymous author translated the last great catechism produced by the Jesuits in the colonial period (by Father João Felipe Bettendorf, published in 1687) into the "vulgar" language, entitled: *Doutrina christaã em lingoa geral dos índios do estado do Brasil e Maranhão, composta pelo P. Philippe Bettendorf, traduzida em lingoa g[eral] irregular, e vulgar uzada nestes tempos* 'Christian doctrine in the general language of the Indians of the State of Brazil and Maranhão, composed by Father Philippe Bettendorf, translated into the general irregular and common language currently used'.

Excerpts of the Catecismo na língua brasílica 'Catechism in the Brazilian Language', were entwined with excerpts from the *Doutrina christaã em lingoa geral irregular e vulgar* 'Christian doctrine in the general irregular and common language', from the 18th century, by Edelweiss (1969, 139–158), with the objective of outlining the changes in standard speech of the Tupi, as separate for the language of the mestiça population. The following table illustrates one part of this comparison:

Table 1: Comparison of excerpts from *Catecismo Brasílico*, by Father Antônio de Araújo (1618), and *Doutrina em lingoa geral vulgar*, anonymous (no date, 18th century).

English	*CATECISMO BRASÍLICO* by Father Antônio de Araújo	*DOUTRINA EM LINGOA GERAL VULGAR* Anonymous
P: Who, then, did God create first to inhabit the earth? R: The one who was to become our first father.	P: *Abá pe erimbaé Tupã oi monhang ypy ybypóramo*? R: *Asé rubypyrama*.	P: *Abá pe erimbaé Tupana o monhang ypy ybypora rama*? R: *Îande paîypy, ou: îandé ramũîypy*
P: What made him into a body? R: Just clay.	P: *Mbaé pe oî monhang setéramo*? R: *Ybyuuma nho*.	P: *Mbaé taé oî monhang setèrama*? or: *mbaé suí* or *monhang seté* ? R. *Yby piranga; ou: yby piranga suí.*

In his analysis, Edelweiss identifies not only alterations of lexical forms, but above all syntactic alterations, such as the loss of the direct object *i* of the verb *monhang* and the loss of adjectival meaning of the names. These examples demonstrate that, although the general languages are scarcely attested in written registers, the sources mentioned allow us to affirm that there were clear differences between the language used by *mestiços*, the Indigenous population, and colonists in the contexts of communicative immediacy and the language described by the Jesuits in grammars and dictionaries for the purposes of the catechism.

This scenario of Portuguese usage by a small elite, especially in the spheres of administration, religion and law, and the use of general languages by the Indigenous population, Africans and settlers in situations of immediacy, is the basis for the emergence of a situation of *diglossia,* which would develop in the centuries that followed between the erudite variants of BP, those of urban and literate origins, and the popular variants of a rural and predominately illiterate one, used by uneducated speakers and pertaining to the lowest levels of the Brazilian social structure.

The beginnings of colonization, then, can be seen as a period of many modifications to the universes of discourse in daily life, religion, science and art in Brazil, through the introduction and adaptation of discourse traditions such as the catechism, works for the theatre, and grammars, among many others, to the linguistic and cultural reality of the New World. However, it is above all in the broad area of religious discourse that the greatest modifications occurred.

3 Oral and written texts in Colonial and Imperial Brazil

From 1690, and particularly in the 18th century, the general languages, which were the language of missionaries, settlers, Indigenous populations and *Mameluks*, lost predominance, with Portuguese increasingly used as a language for general use within Brazilian society (Faraco 2016, 141; Lucchesi 2009, 47–52).

On the one hand, in 1757, the administration of the Marquis of Pombal, a representant of the Enlightenment (*iluminista*), edited the *Directorio que se deve observar nas povoaçõens dos índios do Pará e Maranhão em quanto Sua Magestade naõ mandar o contrario* 'Directory that should be observed in the populations of the Indians of Pará and Maranhão as long as His Majesty does not order the contrary', abolishing the use of the general languages and imposing the use of Portuguese in the colonies. In addition, by expelling the Jesuits from Brazil, Pombal's policy obstructed the main discursive-traditional channel (the religious discourse) that had provided the contact of the speakers of the popular variety with the communicative distant varieties. This had a disastrous impact on teaching in Brazil, which only began to be reversed when the Royal Portuguese family arrived (1808).

On the other hand, the African population and their descendants played a decisive role in the dynamics of multilingualism in colonial and imperial Brazil. Between 1538 and 1855 millions of Africans were brought forcibly to the country as slaves. In the sugar producing zones, the Indigenous presence diminished progressively, while the number of African slaves increased. This mass of enslaved African people, on arriving in Brazil, had to learn, orally and without any school-based standardization, the language of their masters, and this from precarious and often impoverished situations of exposure to the target language; thus the general languages ceded space to Portuguese, the language spoken between the enslaved and the settlers.

Within this timeframe—from the mid-16th century to the end of the 19th century—emerged the conditions that led to the situation of the diglossia of BP, with the diffusion of erudite varieties by a small literate elite, and popular varieties by Indigenous people, Africans and *mestiços*, who constituted the majority of the population and who acquired and used the language in situations of communicative immediacy and of irregular linguistic transmission.

With respect to literate culture in 18th-century Brazil, Duran (2013, 55ff) describes how the implementation of reading in the conquered lands had been impeded by the censure of lay books by the religious officers of the *Real Mesa Censória* 'Royal Censure Committee', instituted in 1768, which sought to monitor works published in the Kingdom, to the point of punishing those publishing books in Brazil with custodial sentences, suspension of duties, banishment or excommunication. Taking into account the charter of 20th of March, 1720, which prohibited the printing of books in Portuguese America, and that of the 16th of December, 1794, which con-

demned the dispatch of books and papers to Brazil, Duran affirms that the circulation of books had not developed in this era. Clearly, in the 18th century, there was an emergence of literary and scientific academies: the *Academia dos Renascidos* 'the Renaissance Academy' (founded in 1759, in Salvador), the *Academia dos Seletos* 'Academy of the Select' (founded between 1751–1752 in Rio de Janeiro), the *Academia Científica* 'Academy of Science' (created in 1771, in Rio de Janeiro), and the *Sociedade Literária* 'Literary Society' (founded in 1786, in Rio de Janeiro). These were places where the literate could meet to study or debate literature, yet these had a fortuitous and circumstantial character, emerging from the need for people to gather for particular celebrations (Barbosa 2013a, 28). For example, according to Ernest Ebel (apud Duran 2013, 52), an Austrian traveler in 19th century Rio de Janeiro, it was natural for Brazilians not to care about studying or reading because there was no need to do so in the street, at home, or during work or leisure activities: it was common to move through Rio without having to read a single word; books were expensive, libraries were few and far between, printing had been outlawed, and even the Fluminense nobles learned along with slaves the first lessons on how to behave in the world. The literate classes in the Kingdom (above all, the clergy, medical doctors, professors and lawyers) constituted a tiny proportion of the population, and these were indeed able to take advantage, in a limited way, of religious manuals, books and gazettes to which they had access (Barbosa 2013a, 22–23). The books most widely circulated were literary and religious in nature (Villalta 1999 apud Barbosa 2013a, 24). This meant that even in Brazil's elite circles, communication had a predominantly oral character until the beginning of the 19th century. According to impressions by Admiral Arthur Phillip, registered in 1789, Rio de Janeiro was a loud city, where people spoke, sang and prayed at full volume in the public squares:

> 'The circumstances, which in this place most astonish a stranger, and particularly a Protestant, are, the great abundance of images dispersed throughout the city, and the devotion paid to them. They are placed at the corner of almost every street, and are never passed without a respectful salutation; but at night they are constantly surrounded by their respective votaries, who offer up their prayers aloud, and make the air resound in all quarters with the notes of their hymns. The strictness of manners in the inhabitants is not said to be at all equivalent to the warmth of this devotion; but in all countries and climates it is found much easier to perform external acts of reputed piety, than to acquire the internal habits so much more essential' (source: Phillip 1789, 34).

One significant text tradition for understanding the communicative processes of the literate culture of the Colony is the legal-administrative tradition (in intersection with the religious). In Brazil, religious-legal forensic texts drew on the determinations of the *Trent Council* (1545–1563) and subsequently had to be adapted, as the New World demanded specific laws that covered the children of the land, the Indigenous population, and Africans brought by force. It is in this context that sees the emergence of the *As Constituições Primeiras do Arcebispado da Bahia* 'First Constitutions of the Archbishop of Bahia', promulgated in 1707, which served as a first code of ecclesiastical

laws created in the Brazilian colony (Simões/Castilho da Costa 2009, 41). Other notable works in this tradition are the book *Economia Cristã dos Senhores no Governo dos Escravos* 'Christian Economy of the Lords in Governing Slaves' (1700) by the Italian Jesuit Priest Jorge Benci who lived in Brazil, and *O Etíope Resgatado, Empenhado, Sustentado, Corrigido, Instruído, Libertado: discurso teológico-jurídico sobre a libertação dos escravos no Brasil* 'The Captured Ethiopian, Performed, Supported, Corrected, Instructed, Freed: theological-legal discourse on the freeing of the slaves in Brazil' (1756) by Father Manoel Ribeiro Rocha (Simões/Castilho da Costa 2009, 43). From these laws, a series of religious-legal genres was produced in the Colony, including wills, inventories, donation letters, baptismal records, marriage records, and death records, among others.

This material also allows us to survey some of the communicative processes that involved enslaved persons and their owners. Thus, in his *Economia Cristã* (1700) 'Christian Economy', Benci reprimands slave owners for obliging their servants to deliver messages and carry out illicit undertakings. The scandals mentioned here by the author came to his knowledge, above all, through the confessions of these slave owners during lent or other devotions. In addition, Benci argues that it is not only a sin for the slave owners to deny the slaves food, clothes, health care and all of their material needs, but also in not allowing slaves to observe saint days and high holy days. It was the obligation of the slave owner, he claimed, to provide spiritual sustenance, including catechizing his slaves. In case he did not want, or was unable, to teach the doctrine of Christ, the owner should take the slaves to schools, company houses or convents to be taught in their own language. In this respect, another commentary by this author seems to indicate that the obligation for spiritual support was as poorly provided as was material support:

> 'But because the priests, curates and slave owners (that the slave owners are also entrusted with everything that was said about the priest) do not look to do it this way, because they don't teach the Christian doctrine to the slaves, or if they teach them, at best, once a year, and this very rashly; for this there is such a great ignorance of things of God in the slaves of Brazil, who are the largest part of the inhabitants'.[1] (source: Benti, 1700, § II, item 81)

Illiterate culture, driven by Portuguese varieties and African languages spoken by Africans and their descendants, is sparsely attested in written testimonies. Narratives or oral news were among the discourse traditions that made up the daily lives of the enslaved. For example, the slaves responsible for providing public offices with water set out every day to the central public fountains to fill their water jugs, and would

1 "Mas porque os Párocos, Curas e senhores (que aos senhores também compete tudo o que dos Párocos está dito) não procuram haver-se deste modo, porque não ensinam a doutrina Cristã aos servos; ou se lha ensinam, quando muito, uma vez no ano, e isto mui à pressa e de corrida; por isso há tão grande ignorância das coisas de Deus nos escravos do Brasil, que são a maior parte dos seus habitadores" (source: Benti, 1700, § II, item 81).

have received all the news of the city at these meeting places, before returning at night to the prisons, taking such information back with them orally (Barbosa 2013b, 10). Other Black African oral traditions in Brazil included tales, proverbs, anecdotes and predictions, among others. Many Portuguese stories were modified by African elders or nursemaids; for example, many Black women walked from one plantation to another, telling stories to other women, the nannies of white children (Cascudo 2006, 155; Freire 1943, 527–528). This function of storytelling is, according to Cascudo, an African discourse tradition, known as *akpalô*, which thrived in Brazil.

Meanwhile, the voice of the slaves appears indirectly or directly in traditions of ecclesiastical, legal and metalinguistic texts. Faraco (2016, 143–145) cites as an example of the metalinguistic tradition a publication from the 17th century, the *Arte da língua de Angola oferecida a Virgem Senhora N. do Rosario, Mãy & Senhora dos mesmos Pretos* 'Art of the language of Angola offered to the Virgin Our Lady of the Rosary, Mother and Lady of the Blacks themselves' (1697) by Father Pedro Dias, and a manuscript from the 18th century by Antônio da Costa Peixoto, published in 1731, which both provide evidence of a vehicle language. The *Arte da língua de Angola* is a grammar of Kimbundu, a language spoken in Salvador by enslaved people from Angola (↗2 The social history of Brazilian Portuguese). Peixoto's manuscript covers vocabularies, dialogues and phrases necessary for communication between slave owners and slaves.

In addition, documents related to visits to Brazil by the *Holy Office Tribunal* between 1590 and 1780, plus reports and testimonies by black and mixed-race slaves, leave us with clues as to the vernacular and also to daily interactions and religious discourse traditions. In this way, for example, *Primeira Visitação do Santo Oficio às partes do Brasil. Denunciações de Pernambuco* 'First Visit by the Holy Office to parts of Brazil, Denunciation from Pernambuco' (1593–1595) includes registers of divining formulas, used frequently by slaves and *freed slaves* in their daily lives to resolve practical problems, as well as spells and other magic formulas that are preserved to this day, such as spell-breaking verses and the evil eye (Souza 2009, 212–241).

Although very rare, Wissenbach (2002, 105) affirms that documents left by literate Africans and their descendants can be found in the testimonies of witnesses in law suits, where fragments of discourse are written, these constituting clues to the language of the African population in excerpts of dialogues transcribed by the legal authorities. Thus, among the criminal law suits involving slaves and *freed slaves* in São Paulo in the second half of the 19th century, we find the following excerpt (see Wissenbach 1998, 127), in which the enslaved Apolinário is asked during an interrogation why he carried arms:

(2) 'He answered—*Whoever goes on the lamb must go prepared*. Who explained his saying thus—A person who is a fugitive is vulnerable to jaguars and the legendary Captain-of-the-Woods and for this reason, and to avoid being their easy prey he had bought the pistol'. Source: AESP, *A Justiça* versus *Apolinário, escravo de*

Francisco Nogueira 'The Judge versus Apolinário, Slave of Francisco Nogueira', 1872.[2]

In addition to fragments of discourse transcribed in judicial acts, Wissenbach mentions that other documents annexed to law suits provide clues to the writing of enslaved persons. These include items of evidence such as short personal notes, lists of objects (written by slaves with the intention of obtaining possessions, especially when the author was a fugitive), plus prayers and orations written on the inside of amulets and scapulars, worn by men and women for protection. Finally, some letters written by slaves exist, although these were never sent, but rather were used as incriminating evidence.

Other sources of writing by literate slaves and freed slaves are presented in Oliveira (2006), who investigated the documentation of a black fraternity in 18th-century Bahia (*Irmandade de Nossa Senhora da Soledade Amparo dos Desvalidos* 'Fraternity of Our Lady of Soledad Amparo of the Disabled', latterly called the *Sociedade Protetora dos Desvalidos* 'Society of the Protection of the Disabled'), uncovering some 290 written documents written by Africans and their descendants throughout the 19th century. By and large, these are meeting minutes, but also bulletins, project presentations and proposals, newsletters, speeches, among other discourse traditions. In this material, it is also possible to see what, for a contemporary reader, would be characteristics of orality and less prestigious varieties. Given such a context, in the following excerpt (3) we show some of the typical phenomena of the language of immediacy and of popular varieties. Many of these are present in contemporary BP, especially in situations of communicative immediacy, independent of the speaker's education.

(3) Minutes of the *Sociedade Protetora dos Desvalidos* – 'Society for the Protection of the Disabled', Bahia, 1864: Oliveira 2006, 578–582, Doc.17)

	Acta da Seicão do dia 21 de Fevereiro de 1864 Prizidencia do Sen*hor* Socio Manoel Leornado Frenandes abri o dicto Sen*hor* a Seicaõ ao meio dia e fei-	Minutes of the Session of the day 21 of February of 1864 Presidency of Mr. Member Manoel Leornado Frenandes open- ed the said gentleman the session at noon and ma-
5	ta a Chamada na forma do costume a chavom-se prezentes 16 Sen*hor* Sócios é lida acta antreou foi a provada [...] Disse	de the call in the usual manner There were present 16 members is read the previous minute was approved Said

2 "Respondeu—*Quem anda fugido deve andar aprecatado*. Que explicasse o seu dito. —Um indivíduo que anda fugido está sujeito a onças e a Capitão-do-Mato e por isso e para não ser presa fácil deles comprara a garrucha".

45	mais o Sen*hor* Prizidente Os Sen*hor* Ouvirão não estou a qui fazendo a bersurdo, e mandan do meter o projeto em votacao por Sedolas feixadas foi a provada unanimente Com 15 votos numero q*ue* estava prezente e todos fica	more the Mr. President You gentlemen heard not I am doing absurd, and order- ing to put the Project to votation in ballots closed it was approved unanimously with 15 votes number which was present and all be-
50	ndo Sujeitos a murta do artigo 38 dosEs- [p3] tatutos deu hum aparte o Socio Cladio Jozé Gomes q*ue* hera munto gra- nde a multa a vista das do Socios respondeu o Sen*hor* Prizidente q*ue* estava	ing subject to the fine of the article 38 of the Sta- tute gave one aside the member Cladio Jozé Gomes that it was very lar- ge the fine in view of the members replied Mr. President that it was
55	boa e q*ue* havera Ser 5$000 para naõ hav- er farta, [...] [...] mandou o Sen*hor* Prizidente passar a Conpetente acta q*ue* abaxo todos Se asjnarão	well and that it would be 5$000 in order to not there be lacking, [...] sent Mr. President the competente minute which beneath are all signed
90	Manoel Leonardo Fer*nandez* Prizidente [...]	Manoel Leonardo Fernandez President

In these meeting minutes, typical elements of the language of distance can be observed: the use of typical formulaic openings and closings; graphic elements and medial writing; low degree of emotionality and cooperation, among others. On the pole of communicative immediacy, clues can be found to the less controlled register of the speech, such as direct discourse, the use of coordination as a predominant junction strategy (above all the additive conjunction "and"), rhotacism ("farta" instead of "falta", "murta" instead of "multa"), diphthongization ("seição" instead of "sessão", "feixadas" instead of "fechadas"), monophthongization ("munto" instead of "muito"); syncope ("antreou" instead of "anterior"), epenthesis ("abesurdo" instead of "absurdo"), among others.

As an overview of this period, then, we can conclude that the universes of discourse that encompass the main traditions of the language of distance (literature, administration, science) were more accessible to the elite as a cultural and economic group, which in turn explains the scant evidence of the circulation of oral and written texts by the lower social strata, comprising Indigenous peoples, Africans, their descendants and *mestiços*. Moreover, it is in the universe of religious discourse that clues can be found of conceptual orality and indices of the intersection of traditions of the language of distance (Catholic traditions and rituals) with popular black and Indigenous traditions. Generally, however, communication in Brazilian society up to the 19th century, whether in literate or illiterate circles, was mainly based on orality. This changes with the introduction of printing in Brazil (1808) and, consequently, with the increase in the circulation of books and the beginning of the press, among other historic events.

4 Oral and written texts in Brazil in the 19th century, and urbanization

From the 18th century onwards, but especially in the 19th century, a series of sociocultural, political and linguistic reconfigurations triggered the diffusion of varieties of erudite BP. Among these were the settling of urban centers in Minas in the 18th century, the transference of the Portuguese Court to Rio de Janeiro in 1808, the arrival of the press in the same year, the increase in scholarly activity in the 19th century with the creation of technical schools, higher education and the mandatory attendance at elementary schools in villages and cities from 1827 onwards, and the intense urbanization that contributed to the dissemination and standardization of BP (Silva 2004, 131; Basso/Gonçalves 2014, 255–259). Due to the increase in formal education and the arrival of the press, written testimonies from the 19th century are infinitely more abundant than those from previous periods. Many of these testimonies allow for a direct identification of markers of both the language of communicative immediacy and distance. In this period, erudite varieties of BP saw extensive diffusion, not only as literary, legal, administrative and religious languages, but also, from the 19th century onwards, as journalistic language.

With respect to social life in Brazil, around the mid-19th century (↗2 The social history of Brazilian Portuguese), Gilberto Freyre (2013) notes that, whether in the mansions of the sugar mill owners or in the luxorious patriarchs' homes in the city, women of the elite were not accustomed to leaving the house to go to the shops: businesses sent young boys to their homes to sell cuts of fabric, and traveling vendors sold rural products door-to-door. Even writers, such as José de Alencar and Joaquim Manuel de Macedo, had boys at their disposal to make house calls to sell their novels. Girls of the most opulent families were often sent to religious boarding schools, where they learned dancing, embroidery, prayers, French and literature, while others received instruction at home from tutors. Boys were sent to boarding schools at a very early age, where they studied Latin grammar, rhetoric, the French classics, history and geography. At fifteen or sixteen years old, they enrolled in military academies, seminaries, or entered higher education to study law or medicine. During this period, oral poetry duels were common among students in theatres, which served as the main form of entertainment for young boys, along with parties, ceremonies, sermons in the church and street processions. Candido (2006, 159) notes that students in São Paulo in particular began many literary projects in the student fraternity houses (*repúblicas*) of the city towards the end of the 1840s, and in particular the tradition of *bestialogics* in prose and verse, which was a deliberately exaggerated discourse without a logical thread. Religion played an important role in the entertainment of all classes in the rustic population, mainly involving processions and mystery plays (Freyre 2013), whereas the men of the city discussed politics and carried out business transactions in the public square. Education in the home was profoundly Catholic and children

learned the Lord's Prayer, The Apostles' Creed and Holy Mary, Hail Mary and the Catechism. They prayed on waking and going to bed, and before sleep would ask for their parents' blessing. With the arrival of the Portuguese Court, it became necessary to create standards of conduct appropriate for a civilized society, and thus there emerged etiquette manuals that dictated standards of behavior and conduct for the elite both in public and private arenas. These manuals gained their greatest popularity at the beginning of the 20th century. An example of this type of literature is the *Código do bom tom ou regras da civilidade e de bem viver* 'Code of good tone, or rules of civility and good living' (1845) by J.I. Roquette, who instructed his readers to greet people, hold conversations, write invitations, to congratulate and how to behave at parties, balls and society events.

Although writers of the period of Romanticism (among them, José de Alencar and Gonçalves Dias) have defended the *brazilianization* of the written language, there are records of a conservative reaction to this idea in which purists tried to impede what, in the words of José Verissimo, was seen as 'inappropriate intrusions of popular language'[3] (Faraco 2016, 153–155). In the 19th century, the controversy centered on José Alencar, accused of carelessness, and at the turn of the 20th on Rui Barbosa and Ernesto Carneiro Ribeiro and on the drafting of the civil code project (Preti 2005, 84; Leite 2006, 76–84). Only from 1922, with the advent of Modernism, did an appreciation of popular varieties arise. One consequence of this situation of diglossia is the fact that erudite and popular varieties have different, although interrelated, histories. The point of view of the purists, in the context of these linguistic controversies of the 19th century, is illustrated by an excerpt from Araripe Jr. in 1888 on this theme (apud Faraco 2016, 153; Pinto 1978, 234):

> 'But, if I am not mistaken, the thoughts of Mr. João Ribeiro are much more diverse than supposes my contradictor. The observations of the Brazilian philologist [Aldolpho Coelho] are below criticism, not because a Brazilian dialect does not exist, but because in them were confused the *tangled speech of the Africans* and other phenomena of this order with what should truly be considered a new element in the Portuguese language'. Source: Araripe Jr., 1960, 97.[4]

Up to the 19th century, Brazil had an eminently rural profile, in which the norms of the literate urban elite distanced themselves greatly from popular norms, those used by the greater part of the population and influenced by the *tangled speech of the Africans*. In this context, particularly worthy of attention are journalistic genres, in that the press had, in some ways and despite high degrees of illiteracy (↗2 The social history

3 "invasões descabidas da língua popular".

4 "Mas, ou eu me engano, ou o pensamento do Sr. João Ribeiro é muito diverso do que supõe o meu contraditor. As observações do filólogo português [Adolpho Coelho] estão abaixo da crítica, não porque não exista o dialeto brasileiro, mas porque naquelas foram confundidos *o falar atravessado dos africanos* e outros fenômenos desta ordem com o que se deve verdadeiramente considerar elemento novo na língua portuguesa".

of Brazilian Portuguese), a linguistic and cultural impact on trends of standardization among the educated urban variety: newspaper texts, which became an integral part of people's daily lives throughout the 19th century, provided contact with printed materials for their (few) readers, but also for the illiterate population, in that they were often read aloud. In this sense, the newspapers served to educate both readers and listeners through argumentative, news-based and literary texts. According to Lopes/Barbosa (2004, 7), while the presence of books in people's personal effects was relatively small, subscriptions to printed publications were a common practice among the literate classes by the end of the 19th century.

Brazilian newspapers at the beginning of the 19th century typically contained just four pages and the discourse traditions present were not very diverse. The genuinely journalistic ones were the letter from the editor and news, alongside reactions of the public to events in the news and to day-to-day problems, these in letters from readers. Most news items were one sentence in length, consisting of a single speech act. Besides this, some official communication genres were included, such as dispatches, official notices, newsletters, public administration announcements, accords, decrees and official announcements, among others. The literary genre *folhetim* (serialized literature) was a common feature of the era. The final pages of newspapers were dedicated to different types of advertisements: for contacting people, for products and services, concerning fugitive slaves, etc. In the second half of the 19th century and at the beginning of the 20th, the genres of official communication disappeared from newspapers, which stopped playing their initial role as a vehicle for this kind of texts. On the other hand, they became much longer and came to include a greater variety of types of news: police news, commercial developments, events from political and daily life, etc. In addition, news items became longer and more informative. Other discourse traditions that now emerged were reporting, the weather report, wanted ads, last minute telegrams, articles by columnists and correspondents, as well as anecdotes, among others.

In the journalistic sphere, the text genres most commonly examined in diachronic studies of BP from the 19th century have been ads, letters from readers, and letters from the editor (see Berlinck/Guedes 2000; Lopes/Barbosa 2004; Gomes 2007; Zavam 2009; among others). Such a preference can be explained by the characteristics of production of these three text genres in this period. In the first case, the person placing an ad would have dictated the text to a typist (Ilari 2013). In the second, letters from readers were published exactly as they were received. Finally, the letter from the editor represented, to some degree, the sociopolitical position of the newspaper, besides being a space to answer and advise specific readers or groups of readers. Because the language was less controlled than that of newspapers today, which have specialized editorial professionals at their disposal for revision and publicity, letters from reader in the 19th century reveal, in linguistic terms, an affinity with linguistic varieties associated with conceptual orality. For this reason, person, place and temporal deixis frequently occur in these texts. The verb tenses of *narratio*—an area in

which the readers relate events as a means of argumentative support—are predominately of the past (perfect, imperfect, etc.). In virtue of the argumentative function, there are also frequent linguistic phenomena related to the expression of epistemic and deontic modality, such as verbs in the imperative, auxiliary modal verbs, performative formulas, etc.

In this sense, Barbosa (2013a, 55–56) argues that with the introduction of the press there occurred an expansion of the networks of communication of both the language of immediacy and of distance. In the province of Rio de Janeiro, for example, there were two types of newspaper in circulation, those from the province itself and those from further afield; the inverse also occurred, with Rio de Janeiro's newspapers circulating outside the provincial area. This resulted in a circuit of information from the center to the periphery and the periphery to the center. The oratory, journalism, the sociopolitical essay increased to the detriment of the *belles lettres*. The former public text genres were of greater interest to intellectuals, who were influenced by Enlightenment ideas, and who sought to integrate Brazil into the timeless world of reason and science (Candido 2000, 223).

Subsequently, this period can be characterized as one of emergent public genres, exemplified by opinion articles, pamphlets, parliamentary speeches, conversations and debates in politico-literary societies, in masonic lodges, on the streets, and in public squares, etc. (Barbosa 2013a, 56; Candido 2000, 231). Despite high rates of illiteracy among the population, the printed word found distribution through oral practices, such as those in the Reading Halls, places where newspapers could be read aloud to those who did not know how to read (Barbosa 2013a, 77–78). In addition to printed material, Barbosa also mentions manuscript newspapers that circulated widely in the period and which were posted at public locations—on the doors of shops, the walls of houses, and elsewhere—to be copied by hand by readers.

5 Oral and written texts in Brazil from the 20th century on and the age of the media

For Lucchesi (2002, 80), throughout the 20th century, a tendency for change in popular Portuguese—a "bottom up" change—was seen not in the direction of normative standards, but in the direction of an urban educated (or semieducated) standard. In standard educated Portuguese, there is a tendency to move away from the normative educated standard of the European matrix, a change that could be defined as "moving down".

Through the 19th century, and indeed more so into the 20th century, the press contributed to the fact that knowledge of varieties of communicative distance extended to different parts of Brazil, giving some impulse for the trends towards linguistic leveling. It is from the beginning of the 20th century, however, that these tendencies re-

ceived greater force with the universalization of basic education and mass communication (radio, cinema, television and, later on, the internet). These led to the widespread distribution of educated urban varieties, among others, throughout the country, influencing the lowest strata of the population and facilitating the diffusion of changes, both "bottom up" and "top down". In other words, communication processes in Brazil were transformed in the 20th century by virtue of the emergence of new technologies, such as the lithograph, photography, cinema, radio, television and, latterly, digital media.

From 1930, in the period of the 'New State' (*Estado Novo*), the Governor Getúlio Vargas contributed to the popularization of the radio, using it frequently to communicate with the mass population. Between 1932 and 1937, 42 new radio stations began broadcasting in the country, and by 1945 this number had reached 111 (Garcia 1982 apud Barbosa 2013a, 186).

Television arrived in the 1950s, and from the 1960s in particular began to play a decisive role in communicative processes in Brazil. For Barbosa (2013a, 228), an understanding of the communication processes that emerged in Brazil from the 1960s onwards is linked to the adoption of communicative action for the masses; that is, there emerged the notion of an audience perceived as a multitude. Thus, the use of radio as the vehicle par excellence for grassroots groups, and as a tool of mass media, was complemented by television. By 1976 there were already 11,603,000 TV sets in Brazilian homes and businesses (Barbosa 2013a, 305).

In the first years of Brazilian television, the programming was considered elitist and typically involved discourse traditions such as shows that copied the radio format, news, tv dramas, movies made for tv, foreign series, Q&A programs, soap operas (*telenovelas*) and literary adaptations, among others (Neto 2004, 86; Mendes Jr. 2004, 230). The elitist phase of the 1950s was followed by the popular phase and the competition for audiences in the 1960s, in which television programming was basically centered around soap operas (*telenovelas)*, foreign series, and live studio shows. This was followed by the servitude to the military regime, in the 1970s, expansion of the 1980s, and sensationalism in the 1990s (Souza 2004).

In a study on programming by 7 Brazilian television networks during the week of October 14th–20th, 1996, Souza (2004) shows that entertainment was by far the dominant category found in tv programming overall. However, a particular specialization for each station by genre or preferred format was also noted. A majority of children's shows were found on Rede Cultura (31 %) and SBT (25 %), while Globo favored films (24 %), news broadcasts (22 %) and soap operas (18 %). Record and Manchete, in turn, produced religious genres (respectively, 35 % and 20 %), Gazeta produced series and variety shows (16 %), while Bandeirantes produced sports genres (18 %).

In aiming to entertain, television establishes a direct affinity between the popular language and televised language, above all in comedy shows and soap operas. In relation to the style of language in news broadcasts, Rezende (2000) found only 147 different words in the Globo Network's National News and 212 different words used in the

news broadcasts of TJ Brazil and SBT networks; taken together, these two-word counts represented two thirds of all words used in all television news broadcasts studied. In addition, three verbs (*ser*, *estar*, to be; *ter*, to have) comprised 27.3 % of the total use of verbs in JN—National News, and 26.6 % for TJ. This tendency is confirmed in criticisms[5] made by Neuber Fischer for the Press Observatory on April 6th, 2016, claiming that journalists overuse colloquial language on television to such an extent that the dialogue between the news anchors are comparable to "bar talk" and that there are even those who, reporting live, have used the form *o cara* 'the dude' to refer to someone previously mentioned in a report.

An example of the relaxation of the prescribed standard of grammatical rules in BP as it is used on a daily basis is the increased use of the pronoun in subject position and the decrease in the retention of the pronoun in the object position, contrary to European Portuguese, in which the retention of the pronoun is very low in the subject position and very high in the object position (Tarallo 1993; Lucchesi 1998). Discussing this phenomenon, Arden (2015) notes that the avoidance of the pronoun in the object position is considered a linguistic strategy in journalism style manuals (*Manual da Redação da Folha de São Paulo* 2001, 'Editing Manual for the *Folha de São Paulo*'; *Manual de Radiojornalismo Jovem Pan* 1993, 'Radiojournalism Manual for *Jovem Pan*'), as well as constituting a frequently attested phenomenon in formal national television news broadcasts, such as the *National News* of the *Rede Globo* 'Globo Network'.

Another example of this type of change, which is a development relevant to journalistic discourse, is the increasing use of slang in writing. The presence of slang can be identified not just in popular journalism, but also in newspapers targeting a wide audience through national distribution, traditional Brazilian press agencies such as the *O Estado de S. Paulo*, *Folha de S. Paulo*, *Jornal da Tarde*, and prestigious magazines with national distribution, such as *Veja*. The following are examples cited by Preti (2004):

(4) Pitta ajudou **laranjas** a obter lucros. (*O Estado de S. Paulo*, 2/4/97—front page headline)
'Pita helped **oranges** (= scapegoats) acquire profits.'

(5) Gil e Gugelmin não crêem em **urucubaca**. (*O Estado de S. Paulo*, 10/07/97)
'Gil and Gugelmin don't believe in **urucubaca** (= voodoo).'

(6) A **grana** da campanha. (Magazine cover, *Veja São Paulo*, 9–15/9/96)
'The **dough** of the campaign.'

Also, in the area of the transformation of journalistic discourse, Dias (2006) detects a tendency to render the facts relating to news more spectacular in the development of the genre—journalistic news—and to use a series of strategies that aim to provoke

5 http://observatoriodatelevisao.bol.uol.com.br/critica-de-tv/2016/04/jornalistas-abusam-da-linguagem-coloquial-na-televisao, last accessed 18.02.2022.

complicity and involvement in the reader. For this reason, especially in the *lead*, but also in other parts of the compositional structure of a news report, fictional discourse practices can be found, including allusions to poems, omniscient and omnipresent narrator, among others.

According to Barbosa (2013a, 274), the 1990s marked the beginning of the popularization of home computers and the diversification of the use of information technology in a more intensive way, thus beginning the long process of transformation in the communication practices that have their mainstay in the digital world. Data from the Secretary of Social Communication of the President of the Republic, in a survey on the consumer habits of the Brazilian population in terms of the media (2016), involving 15,050 individuals, found the following:

- **Television** is the predominant means of communication, cited by 89 % as first or second means of obtaining information and 77 % watch television every day. Brazilians spend an average of 3 hours and 21 minutes per day exposed to television, from Monday to Friday, and 3 hours and 39 minutes at the weekend.
- **The internet** is the second most frequent means of communication used by Brazilians, cited by 49 % as first or second means of obtaining information, with users connected on average 4 hours and 44 minutes per day during the week, and 4 hours and 32 minutes at the weekend. Users between 16 and 24 years old are connected on average 6 hours and 17 minutes. Among users with up to an 8th grade education, the numbers fall to an average of 4 hours and 4 minutes, and among users with up to a 4th-grade education, to an average of 3 hours and 19 minutes. In the previous edition of the same research (2014), it has shown that among internet users, 92 % are connected through **social networks**, with **Facebook** (83 %), **Whatsapp** (58 %) and **Youtube** (17 %) being the most popular.
- **Radio** is the third most frequent (30 %) means of communication used by Brazilians as the first or second means of obtaining information. Brazilians spend an average of 3 hours and 8 minutes per day exposed to radio, from Monday to Friday, and 2 hours and 44 minutes at weekend. 37 % of people who listen to the radio do so while they are performing another domestic activity, and 17 % listen while they are eating. Also, similarly to television, radio has a social component as a catalyst, since it serves as a background to conversation among people (17 % of listeners).
- Newspapers are cited by 12 % as the first or second mean of obtaining information; magazines are cited by 1 %. The frequency and intensity of the reading of **newspapers** and **magazines** is low. The percentage of Brazilians who read newspapers is 32 %, with an average of 1 hour and 10 minutes per day. 23 % of Brazilians read magazines, with an average of 1 hour and 12 minutes during the week and 1 hour and 14 minutes at the weekend.

On the other hand, the survey *Retratos da Leitura no Brasil* 'Portraits of Reading in Brazil' (Failla [4]2016), which assessed the impact of reading in Brazil, included the fol-

lowing findings based on a sample of 5,012 interviews carried out between November and December 2015:

- Only 23 % of those interviewed bought a book in the 3 months prior to the study;
- 56 % of those interviewed read at least one book in the three months prior to the study;
- In this latter set of readers (2,798), the genres they read most, in order of frequency, were: the Bible; other religious genres; short stories, novels, textbooks, children's books, comic books and role-playing games.

Two especially significant issues emerge from the above surveys. First, a large part of the communicative processes in the daily life of Brazilians involves communicative situations influenced by conceptual orality. Second, the consumption of tv programs and books reveals that religious discourse may be one of the main models of writing for Brazilians. This is currently a poorly investigated issue in the country.

In the sphere of religious discourse, the expansion of popular language to discourse traditions previously restricted to erudite varieties, which present an affinity with the language of communicative distance, can be illustrated by a biblical paraphrase that became an internet phenomenon: the *Bíblia Freestyle* 'Freestyle Bible' (2013), written by the pastor Ariovaldo Carlos Jr., from Minas Gerais, who proposed creating "funny reading of the Sacred Scriptures, interacting with pop culture from the generation Y" (https://www.youtube.com/user/bibliafreestyle/about, last accessed 18.02.2022[6]). Drawing on his experience of evangelism and realizing that conventional biblical language is not easily accessible to everyone, he wrote a paraphrased version of the Bible in the language of heavy metal, punk and other subcultures. Despite its overwhelming public success, the Freestyle Bible generated controversy for its use of slang and profanity, as illustrated in the following table, which includes a passage from Luke 6:10–11 in the *Bíblia Freestyle* and in translation by João Ferreira de Almeida.

6 Igor Jacaúna (2014), *Bíblia Freestyle* (Version 1.0.0.9) [Mobile application software], retrieved from https://play.google.com/store/apps/details?id=br.com.bibliafreestyle&hl=pt_BR&gl=US, last accessed 18.02.2022.

Table 2: Comparison of the passage of Luke 6:10–11 in the *Bíblia Freestyle* and in the translation of João Ferreira de Almeida.

Luke, chapter 6	*Freestyle Bible*	Translation by João Ferreira de Almeida
Verse 10	Diante do silêncio da **galera**, disse: **"Estica** o braço **aí**, **dom"**. E **o cara** esticando a mão, percebeu que tinha sido curado. 'Faced with the silence of the **gang**, he said: '**Stick out** your arm **there**, **chief**'. And the **dude** stuck out his hand, and realized he had been cured.'	E, olhando para **todos em redor**, disse ao homem: **Estende** a tua mão. E **ele** assim o fez, e a mão lhe foi restituída sã como a outra. 'And, looking at **all those around him**, he said to the man: **Reach out** your hand. And so **he** did, and his hand was recovered and seemed just like the other.'
Verse 11	Os crentes ficaram **putos da vida**, e começaram a pensar numa maneira de **se livrarem** de Jesus. 'The believers got **pissed off**, and started thinking of a way to **ditch** Jesus.'	E ficaram **cheios de furor**, e uns com os outros conferenciavam sobre o que **fariam** a Jesus. 'And they became **furious**, and congregated with each other about what they were going to **do** to Jesus.'

So, the relationship between written and oral texts in Brazil, from the second half of the 20th century a top-down change is observed, with the diffusion of norms used by the media, and also a bottom-up change, with an increase in the prestige of popular language. This has occurred above all through a process of cultural uniformization, to which Preti (2005, 23–24) attributes three factors: 1) the democratization of the culture and the increased access of Brazilians to school, university, and digital information; 2) the standardization of leisure activities, in such a way that contemporary media can seek to reach all social classes indiscriminately, influencing them culturally and linguistically, and thus forging common linguistic usage; and 3) new linguistic attitudes —the very fruit of these cultural transformations—which reject the inflexible normative character of the grammatical tradition, and embrace language in use.

6 Final Considerations

In the preceding pages we have presented an overview of some of the main discourse traditions pertaining to the communicative economy of Brazilian society at various stages in time. From the point of view of the immediacy-distance continuum, the history of communication in Brazilian society points to the expansion of BP, motivated by cultural ruptures (political, economic, religious, philosophical, etc.), in the direction of the discourse traditions of the language of distance. At the beginning of the colonial period, the introduction of discourse traditions from the universe of religion through the arrival of Roman Catholicism can be observed. It can be exemplified by

the central role of the catechism and Jesuit works of religious drama (autos). From the colonial period to the beginning of the 19th century, communication within Brazilian society was based predominately on the discourse traditions of communicative immediacy. This situation began to change with the introduction of printing in Brazil, and consequently with the wider diffusion of books and newspapers. In the 19th century, the "public genres" noted by Antonio Candido, such as parliamentary discourse, essays, articles, pamphlets, among others, proliferated as a result of the expansion of journalism, the influence of the ideas of the Enlightenment, and increasing urbanization, thus accelerating the advance of BP in the sphere of the language of distance. From the 20th century onwards, with the universalization of basic education and mass communication (radio, cinema, television and the internet), communication processes in Brazil have been transformed, so that a top-down change can be observed with the dissemination of discourse traditions and norms employed by the media, and also a bottom-up change, with the expansion of popular language to discourse traditions which were formerly restricted to erudite varieties.

7 References

Araripe Jr., Tristão de Alencar (1960), *Obra Crítica de Araripe Júnior: 1888–1894*, Brasília/Rio de Janeiro, Casa de Rui Barbosa.

Arden, Mathias (2015), *Inszenierte Mündlichkeit und elaborierte Mündlichkeit bei TV Globo. Zur soziostilistischen Modellierung morphosyntaktischer Variablen des brasilianischen Portugiesisch*, Frankfurt, Lang.

Barbosa, Marialva (2013a), *História da comunicação no Brasil*, Petrópolis/Rio de Janeiro, Vozes.

Barbosa, Marialva (2013b), *Mundo de práticas orais: processos comunicacionais e matrizes históricas e culturais do público brasileiro*, Revista Conhecimento Online 5/1, 1–16, https://doi.org/10.25112/rco.v1i0.239.

Barros, Maria Cândida Drumond Mendes (1995), *Os "linguas" e a gramática Tupi no Brasil (Século XVI)*, Ameríndia 19/20, 3–14.

Barros, Maria Cândida Drumond Mendes (2003), *A relação entre manuscritos e impressos em tupi como forma de estudo da política linguística jesuítica no século XVIII na Amazônia*, Revista Letras 61, special issue, 125–152.

Barros, Maria Cândida Drumond Mendes/Lessa, Antonio Luis Salim (2005), *Estudo preliminar de um dicionário português-tupi do período pombalino*, Revista de Estudos da Linguagem 13/2, 73–94.

Basso, Renato Miguel/Gonçalves, Rodrigo Tadeu (2014), *História concisa da língua portuguesa*, Petrópolis, Vozes.

Berlinck, Rosane de Andrade/Guedes, Marymarcia (edd.) (2000), *E os preços eram commodos. Anúncios de jornais brasileiros—século XIX,* São Paulo, Humanitas.

Brasil. Presidência da República. Secretaria de Comunicação Social (2014), *Pesquisa brasileira de mídia 2014: hábitos de consumo de mídia pela população brasileira*, Brasília, Secom.

Brasil. Presidência da República. Secretaria Especial de Comunicação Social (2016), *Pesquisa brasileira de mídia 2016: hábitos de consumo de mídia pela população brasileira*, Brasília, Secom.

Candido, Antonio (2000), *Formação da literatura brasileira: momentos decisivos*, Belo Horizonte, Itatiaia.

Candido, Antonio (2006), *Literatura e sociedade*, Rio de Janeiro, Ouro sobre Azul.
Cascudo, Luís da Câmara (1953), *Cinco livros do povo. Introdução ao estudo da novelística no Brasil*, Rio de Janeiro, José Olympio.
Cascudo, Luís da Câmara (2006), *Literatura oral no Brasil*, São Paulo, Glocal.
Castro Jr., João Batista (2011), *Língua Portuguesa, língua tupi e língua geral: jesuítas, colonos e índios em São Paulo de Piratininga; o que entendiam, o que praticavam, o que conversavam*, Doctoral thesis, Salvador, Universidade Federal da Bahia, Instituto de Letras.
Coseriu, Eugenio ([3]1994, [1]1980), *Textlinguistik. Eine Einführung*, Tübingen, Francke.
Daher, Andréa (1998), *Escrita e Conversão: a gramática tupi e os catecismos bilíngües no Brasil do século XVI*, Revista Brasileira de Educação 8, 31–43.
Dias, Ana Rosa Ferreira (2006), *Mídia e interação: estratégias de envolvimento do leitor na construção do discurso da notícia*, in: Dino Preti (ed.), *Oralidade em diferentes discursos*, São Paulo, Humanitas, 111–127.
Duran, Maria Renata da Cruz (2013), *Retórica à moda brasileira: transições da cultura oral para a cultura escrita no ensino fluminense de 1746 a 1834*, São Paulo, Editora Unesp.
Ebel, Ernst (1972), *O Rio de Janeiro e seus arredores em 1824*, Portuguese version by Joaquim de Sousa Leão, São Paulo, Companhia Editora Nacional.
Edelweiss, Frederico (1969), *Estudos tupis e tupi-guaranis*, Rio de Janeiro, Brasiliana.
Failla, Zoara ([4]2016), *Retratos da Leitura no Brasil*, Rio de Janeiro, Sextante/Instituto Pró-Livro.
Faraco, Carlos Aberto (2016), *História sociopolítica da Língua Portuguesa*, São Paulo, Parábola.
Folha de S. Paulo (2001), *Manual da Redação*, São Paulo, Publifolha.
Freyre, Gilberto (2013), *Vida social no Brasil nos meados do século XIX*, São Paulo, Global.
Garcia, Nelson Jahr (1982), *O Estado Novo: ideologia e propaganda política*, São Paulo, Loyola.
Gomes, Valéria Severina (2007), *Traços de mudança e de permanência em editoriais de jornais pernambucanos: da forma ao sentido*, Doctoral Thesis, UFPE.
Ilari, Rodolfo (2013), *Observações preliminares sobre um fragmento do corpus do PHPB: os textos dos jornais de Natal transcritos por Silara*, unpublished.
Instituto Pró-Livro ([4]2016), *Retratos da Leitura no Brasil*, http://plataforma.prolivro.org.br/retratos.php, last accessed 28.02.2022.
Kabatek, Johannes (2006), *Tradições discursivas e mudança lingüística*, in: Tânia Lobo/Ilza Ribeiro/Zenaide Carneiro/Norma Almeida (edd.), *Para a história do português brasileiro*, vol. VI: *Novos dados, novas análises*, tomo II, Salvador, EDUFBA, 505–527.
Kabatek, Johannes (2018), *Lingüística coseriana, lingüística histórica, tradiciones discursivas*, Frankfurt/Madrid, Vervuert/Iberoamericana.
Kloss, Heinz ([2]1978), *Die Entwicklung neuer germanischer Kultursprachen seit 1800*, Düsseldorf, Pädagogischer Verlag Schwann.
Koch, Peter (1997), *Diskurstraditionen: zu ihrem sprachtheoretischen Status und ihrer Dynamik*, in: Barbara Frank/Thomas Haye/Doris Tophinke (edd.), *Gattungen mittelalterlicher Schriftlichkeit*, Tübingen, Narr, 43–79.
Koch, Peter/Oesterreicher, Wulf ([2]2011), *Gesprochene Sprache in der Romania: Französisch, Italienisch, Spanisch*, Berlin/New York, De Gruyter.
Leite, Marli Quadros (2006), *Metalinguagem e discurso: a configuração do purismo brasileiro*, São Paulo, Humanitas.
Lopes, Célia Regina dos Santos/Barbosa, Afrânio G. (edd.) (2004), *Críticas, queixumes e bajulações na imprensa brasileira do século XIX: cartas de leitores e cartas de redatores*, Rio de Janeiro/São Paulo, Humanitas.
Lucchesi, Dante (1998), *A constituição histórica do Português Brasileiro como um processo bipolarizador: tendências atuais de mudança nas normas culta e popular*, in: Sybille Große/Klaus Zimmermann (edd.), *Substandard e mudança no Português do Brasil*, Frankfurt, TFM, 73–99.

Lucchesi, Dante (2002), *Norma linguística e realidade social*, in: Marcos Bagno (ed.), *Linguística da Norma*, São Paulo, Loyola, 63–92.
Lucchesi, Dante (2009), *História do contato entre línguas no Brasil*, in: Dante Lucchesi/Alan Baxter/Ilza Ribeiro (edd.), *O português afro-brasileiro*, Salvador, EDUFBA, 41–73.
Mendes Jr., Osmar (2004), *Antenas paulistanas: imagens em branco e preto*, in: José Marques de Melo/Antonio Adami (edd.), *São Paulo na Idade Mídia*, São Paulo, Arte e Ciência, 225–244.
Neto, Samuel Pfromm (2004), *Biografia de um jornal do interior paulista*, in: José Marques de Melo/Antonio Adami (edd.), *São Paulo na Idade Mídia*, São Paulo, Arte e Ciência, 71–92.
Oliveira, Klebson (2006), *Negros e escrita no Brasil do século XIX: sócio-história, edição semidiplomática de documentos e estudo lingüístico*, Doctoral Thesis, Salvador, Universidade Federal da Bahia.
Phillip, Arthur (1789), *The Voyage of Governor Phillip to Botany Bay with an Account of the Establishment of the colonies of Port Jackson and Norfolk Island; compiled from Authentic Papers, which have been obtained from the several Departments to which are added the Journals of Lieuts. Shortland, Watts, Ball and Capt. Marshall with an Account of their New Discoveries, embellished with fifty five Copper Plates, the Maps and Charts taken from Actual Surveys, and the plans and views drawn on the spot, by Capt. Hunter, Lieuts. Shortland, Watts, Dawes, Bradley, Capt. Marshall, etc*., London, Printed for John Stockdale, Piccadilly.
Pinto, Edith Pimentel (1978), *O português do Brasil: textos críticos e teóricos*, vol. 1: *1820–1920*, São Paulo/Rio de Janeiro, EDUSP/Livros técnicos e científicos.
Porchat, Maria Elisa (1993), *Manual de radiojornalismo Jovem Pan*, São Paulo, Ática.
Preti, Dino (2004), *Transformação no fenômeno sociolinguístico da gíria*, in: Dino Preti, *Estudos de língua oral e escrita*, Rio de Janeiro, Lucerna, 99–107.
Preti, Dino (2005), *A propósito do conceito de discurso urbano oral culto: a língua e as transformações sociais*, in: Dino Preti et al. (edd.), *O discurso oral culto*, São Paulo, Humanitas, 21–34.
Preti, Dino, et al. (edd.) (2005), *O discurso oral culto*, São Paulo, Humanitas.
Rezende, Guilherme Jorge (2000), *Telejornalismo no Brasil*, São Paulo, Summus.
Rodrigues, João Barbosa (1890), *Poranduba amazonense, ou kochiyma-uara porandub*, Rio de Janeiro, Typ. de G. Leuzinger & Filhos.
Silva, Rosa Virgínia Mattos e (2004), Ensaios para uma sócio-história do Português Brasileiro, São Paulo, Parábola.
Simões, José da Silva/Castilho da Costa, Alessandra (2009), *As atas paroquiais de batismo, casamento e óbito como gêneros discursivos*, in: Maria Silvia C. Beozzo Bassanezi/Tarcísio R. Botelho (edd.), *Linhas e entrelinhas: as diferentes leituras das atas paroquiais dos setecentos e oitocentos*, Belo Horizonte, Veredas & Cenários, 35–58.
Souza, José Carlos Aronchi de (2004), *Gêneros e formatos na televisão brasileira*, São Paulo, Summus.
Souza, Laura de Mello e (2009), *O diabo e a terra de Santa Cruz: feitiçaria e religiosidade popular no Brasil colonial*, São Paulo, Companhia das Letras.
Stahlhoefer, Alexander De Bona (2015), *Bíblia Freestyle: oportunidades e dificuldade na contextualização do texto bíblico*, Azusa: Revista de Estudos Pentecostais 4, 1–41.
Steinen, Karl von den (1894), *Unter den naturvölkern Zentral-Brasiliens. Reiseschilderung und Ergebnisse der zweiten Schingú-Expedition, 1887–1888*, Berlin, Geographische Verlagsbuchhandlung von Dietrich Reimer.
Stradelli, Ermano (1929), *Vocabuloários da Língua Geral Português-Nheengatu e Nheengatu-Português, Precedidos de um Esboço de Gramática Nheenga-Umbuesaua Miri e Seguidos de Contos em Língua Geral Nheengatu Poranduba*, in: Revista do Instituto Histórico e Geográfico Brasileiro 104/158, Rio de Janeiro, Imprensa Nacional, 9–768.

Tarallo, Fernando (1993), *Diagnosticando uma gramática brasileira: o português d'aquém e d'além-mar ao final do século XIX*, in: Ian Roberts/Mary Aizawa Kato (edd.), *Português Brasileiro: uma viagem diacrônica*, Campinas, Ed. da UNICAMP, 69–106.

Villalta, Luiz Carlos (1999), *Os leitores e os usos dos livros na América Portuguesa*, in: Márcia Abreu (ed.), *Leitura, história e história da leitura*, Campinas/São Paulo, Mercado de Letras/Fapesp, 90–105.

Wissenbach, Maria Cristina Cortez (1998), *Sonhos africanos, vivências ladinas*, São Paulo, Hucitec.

Wissenbach, Maria Cristina Cortez (2002), *Cartas, procurações, escapulários e patuás: os múltiplos significados da escrita entre escravos e forros na sociedade oitocentista brasileira*, Revista Brasileira de História da Educação 4, 103–122.

Zavam, Aurea Suely (2009), *Por uma abordagem diacrônica dos gêneros do discurso à luz da concepção de tradição discursiva: um estudo com editoriais de jornal*, Doctoral thesis, Fortaleza, Universidade Federal do Ceará.

Patricia Carvalhinhos

18 Onomastics and toponomastics

Abstract: This chapter offers a general overview of the history of proper names in Brazil from the colonial period until current times: place names and personal names. This historical overview focuses on the contact between Native groups and Europeans, emphasizing the onomastic consequences of these contacts. From a structural perspective, the chapter briefly describes the main features of Brazilian toponyms and the primary constituents found in Native Brazilian place names, and illustrates some aspects of ethnonyms and glossonyms. We also include examples of urban toponymy. Finally, the chapter provides insights into certain aspects of Brazilian anthroponymy.

Keywords: proper names, toponymic grammar, toponym structure, hodonyms, Tupi-based names, anthroponyms, constituent elements of naming

1 Opening observations

Scholars of Brazilian toponomastic studies use data and cartographical resources produced by the Instituto Brasileiro de Geografia e Estatística, 'Brazilian Institute of Geography and Statistics'—IBGE, an influential federal organization whose function is to produce, analyze and provide statistical, cartographic and geographic data. Despite its importance, the institute has produced few studies and programs dedicated to onomastics itself.[1] Hence, in Brazil the most commonly used terminology is that proposed by the United Nations Group of Experts on Geographical Names (UNGEGN). In agreement with the International Council of Onomastic Sciences (ICOS), "onomastics [is] the study of proper names in a scholarly way." In this context, it is sometimes necessary to consider transfer phenomena between common and proper names. One of the concepts related to such phenomena is *onymization*, the "[...] transfer of a linguistic unit (including common nouns, adjectives, verbs, interjections, phrases, etc.) to the class of proper names" (ICOS 2021, 4). The main branches of onomastics are *toponomastics*, the study of place names, and *anthroponomastics*, the study of personal names. According to Peters (2014, 4), who brilliantly synthesizes in a few paragraphs Nash/Simpson's study (2012),

> 'How a society names its surroundings exposes how people conceptualize their world: how landforms are distinguished, how people interact with the land, how land is owned or distributed, or how places or topographic features are important spiritually, politically, or historically [...]. In a

1 Nevertheless, the IBGE is an institute of great importance to support Brazilian scholars in terms of providing data and cartography, and will be mentioned several times in this chapter.

https://doi.org/10.1515/9783110405958-019

> toponym is encoded a literal meaning, a lexified denotation, historical and folk etymologies, connotations, and physical properties of the place described [...]. In addition, toponyms may preserve linguistic features that do not appear elsewhere in the language with any frequency, particularly locative morphemes [...]. For these reasons, toponyms can comprise an important portion of linguistic data.' (All translations from Portuguese are my own.)

Toponomastics incorporates the study of all place names, including proper names for cultural and natural places like watercourses and mountains. Toponomastics and anthroponomastics are independent areas of scientific study.

2 Pre-colonial and colonial toponymy

2.1 Pre-colonial epoch and the supposed names

At the time of the Europeans' arrival in Brazil, Native groups had not yet developed a writing system. Therefore, their written place names are in part reconstructed. The first toponyms for which we have references date from the initial period of contact between Europeans and Native peoples. In some cases, it might be that the first Europeans had interpreted referential or locative terms in speech as proper names. In this sense, we do not know if there was a functional toponymic system in Brazilian prehistory. Brazilian place names only began to be registered in the 16th century by the Portuguese and other Europeans (French, Spanish, Dutch). All 16th century Native toponyms are known by the reports of the first Europeans to arrive, and later by colonizers' reports.

2.2 The colonial toponymy: Native (or traditional) peoples in the 16th century

The voyages to Brazil in the 16th century provided a fantastic experience for Europeans: to come into contact with a completely different lifestyle and conception of the world. Native peoples from the coast spoke a common and rather homogeneous language, *the most spoken language along the Brazilian coast* (see ↗1 Brazilian Portuguese linguistics: an overview, and ↗2 The social history of Brazilian Portuguese). At the time, Europeans were discovering a new and utterly different reality, including flora and fauna. They could not have imagined the diversity of the Native peoples that lived in the territory. The Native tribes of the interior were called Tapuia, 'strangers', 'outlanders' by those in the coastal areas. According to Monteiro (2009, 19),[2]

2 "Ao chegarem ao Brasil, os invasores europeus logo descobriram que grande parte do litoral bem como as partes do interior às quais se tinha mais acesso encontravam-se ocupadas por sociedades que

> 'Upon arriving in Brazil, the European invaders quickly discovered that most of the coastal regions and parts of the upcountry regions with better access were occupied by societies that shared some basic features common to the so-called Tupi-Guarani culture. Despite the presumed uniformity, any attempt to synthesize Brazil's sixteenth-century ethnographic conditions stumbles on two problems. First of all, the Tupi society remained radically segmented, and relations between sections or even between local units often boiled down to military actions. [...] Second, non-Tupi societies also inhabited most parts of Brazil, representing dozens of extinguished language families'

The land was gradually being conquered, from the coastal region to the hinterland. The invaders enslaved Amerindians to provide a workforce and identified them as their friends or enemies. Linguistically, the upcountry Tapuia spoke a language different from that of the coastal people, which resulted in misinterpretations and the term *índios de língua travada*, something like 'tongue-tied Natives'. Monteiro (2009, 19) calls this an ethnological simplification:

> 'To tackle these problems, sixteenth-century Europeans tried to reduce the vast ethnographic panorama into two general categories: Tupi and Tapuia. The Tupi part of this dichotomy encompassing the coastal societies in close contact with the Portuguese, French, and Spaniards, from Maranhão to Santa Catarina, including the Guarani people. If it is true that these groups showed similarities between their traditions and cultural patterns, the same cannot be said of the Tapuia. Actually, Europeans usually apply the name 'Tapuia' to groups that were little known to them. Besides, the Tapuia standards used to be socially different from Tupi.'[3]

This section takes an onomastic perspective to address the 16th-century dialectological situation and further linguistic features analyzed in other chapters (↗1 Brazilian Portuguese linguistics: an overview; ↗2 The social history of Brazilian Portuguese; ↗3 The history of linguistic contact underlying Brazilian Portuguese). Place names, when considered as an ethnographic object, demand to be understood in terms of dialectology for their correct analysis.

---

compartilhavam certas características básicas, comuns à chamada cultura tupi-guarani. Contudo, a despeito das aparências de homogeneidade, qualquer tentativa de síntese da situação etnográfica do Brasil quinhentista esbarra imediatamente em dois problemas. Em primeiro lugar, a sociedade tupi permanecia radicalmente fragmentada, sendo que as relações entre os segmentos ou mesmo entre unidades locais frequentemente resumiam-se a ações bélicas. [...] Em segundo, grande parte do Brasil também era habitada por sociedades não tupi, representando dezenas de famílias linguísticas distintas."

3 "Para enfrentar estes problemas, os europeus do século XVI procuraram reduzir o vasto panorama etnográfico a duas categorias genéricas: Tupi e Tapuia. A parte tupi desta dicotomia englobava basicamente as sociedades litorâneas em contato direto com os portugueses, franceses e castelhanos, desde o Maranhão a Santa Catarina, incluindo os Guarani. Se é verdade que estes grupos exibiam semelhanças nas suas tradições e padrões culturais, o mesmo não se pode afirmar dos chamados Tapuia. De fato, a denominação 'Tapuia' aplicava-se frequentemente a grupos que—além de diferenciados socialmente do padrão tupi—eram pouco conhecidos dos europeus."

2.3 The toponym Brazil

Brazil had several concurrent names during the exploration stages of its territory. In a letter by Pero Vaz de Caminha to the King of Portugal, Dom Manuel I, written on May 1st, 1500, the land was referred to as *Ilha de Vera Cruz* 'True Cross Island' (↗7 The history of the lexicon). There is no evidence of Native denominations in this document. Systematic records appeared over eighty years later, with Gabriel Soares de Sousa's *Tratado Descritivo do Brasil* 'Descriptive Treaty of Brazil' (1587). Cartographers had produced several maps during the first decades following the use of the names of Vera Cruz/Santa Cruz/Brazil. The cartographical record shows that brazilwood featured on numerous South America maps that appeared before the name Brazil itself (Rocha/Presotto/Cavalheiro 2007, 754 and 761–762):

> 'The brazilwood represented in many maps of South America probably gave the official name to the Brazilian territory, which for the first time was Ilha de Vera Cruz, then Terra de Santa Cruz, and finally Brazil.
>
> The first map which used the name 'Brazil' to indicate Terra de Santa Cruz was the Planisphere Orbis Typus Universalis Tabula by Jerônimo Marini (1512) made in Venice, Italy (Matos 1999). However, the Mapa de Cantino (1502) already used the name Brazil to identify a river, where the expedition (1501/1502) commanded by Gonçalo Coelho found assortments of brazilwood.'

As reported by Rocha/Presotto/Cavalheiro (2007), several maps, drawn by cartographers from different parts of the world, use other exonyms to refer to Brazil, like Waldseemüller, who called it *Terra Papagalli* 'Land of Parrots':

> 'Martin Waldseemüller (1470–1521), in a map published in the edition of *Ptolomeu's Geography* (1513), and republished by Lorenz Fries (approximately 1490–1532) in 1541, designates Brazil as *Terra Papagalli* (Land of Parrots), and used the term Brazil only to name the river located between Porto Seguro and Monte Pascoal, in the actual Bahia State.' (Rocha/Presotto/ Cavalheiro 2007, 763–764).

However, in Pero Magalhães Gândavo's laudatory *Historia da prouincia sa(n)cta Cruz a qui' vulgarme(n)te chamam Brasil* 'History of the province of Santa Cruz',[4] the title itself proposes two new place names that emerged over the first seventy-odd years of history of the country. The first was *Terra de Santa Cruz* or 'The Land of the Holy Cross', now recognized as a land, and not as an island (an error corrected during the first landholdings); the second, finally, was *Brazil*. In many other places, the religious motif is found in those early days.

4 Published in 1576, this is a portrait of the Native peoples and the fauna and flora of Brazil, written in a marketable style. It is the first book about the Brazilian lands, and was published in Lisbon.

The records of the first place names in Brazil, as Dick observes (1990, 315–316), appear in geographic charts drawn up by Europeans such as Alberto Cantino and Nicholas Canério in the first decade of the 16th century. They are seen in Duarte Pacheco Pereira's book (15??, 1892), *Tábua de ladezas ou latitudes do Esmeraldo de Situ Orbis*.[5] This work registered toponyms such as Ilha da Ascenção and Ilha de Santa Clara (*ilha* 'island'), Aguada de São Miguel (*aguada* 'place where ships can obtain supplies of drinking water'), Rio de São Francisco (*rio* 'river'), Angra de Todos os Santos (*angra* 'cove'), Angra de São Roque, Ilha de São Lourenço and Ilha de Santo Amaro, Cabo de Santo Agostinho (*cabo* 'cape'), Rio de Santa Luzia, Ilha de Santa Bárbara, and Ilha de Santa Maria da Arrábida.

The commercial activity surrounding brazilwood (*Caesalpinia echinate*, Tupi *ybyr-apitanga*) led to the toponym that the country still carries. The etymology of *Brazil* is often seen as controversial. According to popular sources, there was a convergence with the term *brasa* 'ember' due to the wood's red color. Murault (2006) conducts extensive and detailed research on inkwoods in several countries, clarifying conflicting lexical situations caused by overlapping references. She shows that Asian inkwood species had been very well known in Europe since the 11th century. Brazilwood is an inkwood species with multiple applications, used for dyeing, marquetry, and lutherie. The author also cites reports from Marco Polo, in 1260, who described some Asian inkwood species. Finally, she notes the terminological diversity of the term *brazil*, applied in the 16th century to *Caesalpinia*, already widely known and widespread in medieval Europe:

> 'In various customs duties, [*brazil*] is among the many drugs and spices imported from the East. The Venetians called it *verçí*, *verzi*, and *verzino* (from the Arab *wars*), a word that has gradually turned into *berzil*, *berzi*, *brezilh*, *bresil*, *bresilium*, *brexilium*, *braxilis*, *brazil* to become *Bresilien-*, *Pressilgen-* or *Prissiligholz* in German and *brazilwood* in English. Thus, Marco Polo named the tree from the dye name given to it, then, the term *brazil* is applied depending on the context, either to the tree or to the dye' (Murault 2006, 174; author's highlights).[6]

5 Duarte Pacheco Pereira's work was never published while he was alive. Its first publication took place in the 18th century. Souza (2011, 92–97) analyzes the work and opinion of several authors who previously worked on this issue, such as Carvalho (1965), a study we also consulted. The author provides details of the history of *Esmeraldo du Situ Orbis*, commissioned by King D. Manuel in 1505, when Duarte Pacheco Pereira probably began writing it. This treatise on geography/cosmography/navigation was intended as a practical manual for navigators, a limited audience. Besides, the Kingdom's most significant interest at the time would have been the publication of works of a religious nature, which might explain why Pacheco Pereira's study was not printed during the author's lifetime. Its publication in the 18th century was based on this incomplete 16th-century manuscript.

6 "Em diversas tarifas aduaneiras, figura entre as incontáveis drogas e especiarias importadas do Oriente. Os venezianos chamavam-na *verçí*, *verzi* e *verzino* (do árabe *wars*), vocábulo que se transformou progressivamente em *berzil*, *berzi*, *brezilh*, *bresil*, *bresilium*, *brexilium*, *braxilis*, *brazil* para se tornar *Bresilien-*, *Pressilgen-*, ou *Prissiligholz* em alemão e *brazilwood* em inglês. Assim, e tendo Marco Polo bati-

The reason for the spontaneous creation of such a trade name was its wide commercial appeal. According to mercantilist thinking, this was enough to ignore the Native name *Pindorama*, a mythical toponym.

It is claimed by some that among the Ando-Peruvian Natives the land we know as Brazil was called *Pindorama* 'land of palm trees'. Others argue that Pindorama was the general American name of Brazil, as expressed by the greetings of Couto de Magalhães, for whom it was also used by the coastal peoples that spoke Tupi or Guarani:

> 'They knew what we now call Brazil, from the Amazon to roughly the Baía dos Patos ['Bay of Ducks'] under the name Pindorama, which means a region of palms; for the upcountry, without palms, they used the term *Tapuirama*, which means an area of ranches or villages'[7] (Couto de Magalhães 1897, 4).

3 Contact languages and ideological clashes

The cartographical record of Brazil in the first decades of the 16th century registers only Portuguese religious toponyms. These place names demonstrate a naming mechanism based on the medieval God-centered world view and the importance of hagiology. Such identification of spaces with toponyms in Portuguese provided Europeans with the ownership of the lands in question, disregarding the original Native toponyms, to please the Catholic Church.

The contact between languages created bilingual speakers and also new denominations (↗3 The history of linguistic contact underlying Brazilian Portuguese). These new names sometimes caused what we might call misinterpretations. Despite learning the land's language, the colonizers did not always translate the toponyms satisfactorily into Portuguese.

The agglutinating nature of Brazilian Native languages made it difficult to perceive the constituents of the toponyms. Even today, the fact that Native languages are agglutinative causes the perception, for speakers of Portuguese, of a single linguistic unit where there are in fact several units. Europeans learning the Native languages sometimes translated common names as proper names.

Some of these terms are given here as examples, followed by toponyms constituted from those same geographic features: YBYTYRA 'hills', in *Ibitira*, BA;[8] YBYTYRA UN

---

zado a árvore a partir do nome do corante por ela fornecido, o termo *brasil* aplica-se, conforme o contexto, ora à árvore ora à matéria tintorial."

7 "Estes conhecião o que chamamos hoje Brazil, do Amazonas até mais ou menos a bahia dos Patos, debaixo do nome de Pindorama, que quer dizer Região das Palmeiras; ao interior, não occupado por elles, denominavão *Tapuirama*, que quer dizer região de ranchos ou de aldeas."

8 The various examples presented are located within their home states, indicated by the official acronyms: Acre (AC), Alagoas (AL), Amapá (AP), Amazonas (AM), Bahia (BA), Ceará (CE), Espírito Santo

A 'dark hill', in *Ibituruna*, MG; PARANÃ 'river', in *Rio Paraná* 'Paraná River', MA, GO, SP; ITÁ-Ī 'small stones', in *Ribeirão Itaim* 'Itaim Creek', SP; and *Rio Itaim* 'Itaim River', MG, PI.

This second stage of shaping place names in Brazil highlights the coexistence of two naming strategies. We can note this when observing the same landscape element named descriptively by Natives and with Europeans' value judgments. These are cases of bi-nomination, or places that have both a Native name and a Portuguese one: *Itacoatiara* (ITÁ KÛA TIAR A 'painted rocks'), *Ilha da Cananea* (nowadays *Ilha de Cananéia* 'Cananéia Island', SP). According to the history of Cananeia,[9] the modern city's name has its origin in the colonial toponym. The old town, founded in 1601, had been named by the Europeans. *São João de Cananéia*, the old place name, comes from Canaan, the promised land of the Hebrews, a biblical reference.

Other examples of this stage are *Pará* 'big river', *Rio de São Francisco da Banda do Norte* (something like 'São Francisco river from the north side'); *Iggoaçû* 'big river',[10] *Rio da Prata* 'silver river', *Jurumiri* 'little mouth', *Rio dos Patos* 'ducks' river'.

Explorers also spread the *língua geral*, the 'general language' (the Church's language to catechize the Natives, see ↗2 The social history of Brazilian Portuguese); meanwhile, the settlers traveled around the inland areas.[11] The exploration of the land and the inclusion of new cultural elements also provided many neologisms appended to the general language, either created by Europeans or by Native peoples.

Since European settlers introduced technologies previously unknown to Native populations, they had to use previously known terms to create semantic neologisms and make themselves understood, but this does not imply that all Brazilian toponyms come from this period. Dick (1990) cites a few examples of neologisms introduced in the general language, such as *itaîuba* 'yellow stone' for 'gold', *itaeté* 'excellent stone, steel', *itamembeka* 'soft stone' for 'lead ore'; in addition to those introduced by the Catholic Church: *îande-îara* 'Our Lord', and *tupã sy* 'Tupã's mother', for 'Our Lady', among others.

The missionaries used the Tupi terms to express the whole range of related meanings, as with *tupã*, whose first meaning is 'entity which makes the rain or thunder fall' (Navarro 2013, 482–483). This word also contemplates the divine character. Thus, the term *tupã* is used in composition with other terms for precise meanings within the

(ES), Goiás (GO), Maranhão (MA), Mato Grosso (MT), Mato Grosso do Sul (MS), Minas Gerais (MG), Pará (PA), Paraíba (PB), Paraná (PR), Pernambuco (PE), Piauí (PI), Rio de Janeiro (RJ), Rio Grande do Norte (RN), Rio Grande do Sul (RS), Rondônia (RO), Roraima (RR), Santa Catarina (SC), São Paulo (SP), Sergipe (SE) and Tocantins (TO). Furthermore, Distrito Federal (DF).

9 IBGE, http://biblioteca.ibge.gov.br/index.php/biblioteca-catalogo?view=detalhes&id=32076, last accessed 18.02.2022.

10 Dick (1990) uses this record according to the *Vocabulário na Língua Brasílica*.

11 Political nit-picking, dissension between government/Church and other intentions of the Jesuits are not discussed in this chapter.

Catholic practice such as *tupãmongetasaba* 'prayer instrument', for *prayer book*; *tupãerogûatá* 'walk with God', for *procession*; *tupãoka* 'House of God', for Church, among others. These and other examples were collected from the *Vocabulário na Língua Brasílica*, an anonymous 17th-century manuscript first published in 1938 (Ayrosa 1938). Hence, determining whether a Native toponym is "Tupi or non-Tupi"[12] is a difficult task. It is difficult to know if the toponym comes from "classic" Tupi spoken before the arrival of the Europeans or whether it was born in the colonial period due to an exchange of experiences between peoples. In general, the coastal toponyms may be considered older and coined directly by the Native peoples, as noted in some coastal cities.

The city of São Paulo, for instance, arose early in the second half of the 16th century, in 1554, and its well-preserved documentation[13] provides records of toponyms applied to the rivers and paths used before the arrival of Europeans. The use of the general language probably led to the creation of the other toponyms that appear throughout the 17th and 18th centuries—place names used not only by Natives but by anyone fluent in the general language, commonly spoken until the second half of the 18th century.

3.1 Modified toponyms and European influences

When removed from their original place of use, the modified toponyms denote namespaces in other lands. In Brazil's colonial times, it is through modified names that Native toponymy suffered ideological attacks. At various moments in Brazil's history, Europeans had the clear intention of erasing Native cultures. However, it is in the second half of the 18th century, under the pressure of Pombal's reforms, that this is most evident.

With a policy which became overwhelming in its influence, several Native place names disappeared and were replaced by Portuguese toponyms (usually legitimated by Royal Charters). These place names are entirely disconnected from the Brazilian landscape and refer to nothing other than the memory of political despotism. Cardoso (1961, 283) quotes some of the replacements related to so-called homesickness:

12 Although not intentional, this expression may recall Oswald de Andrade's *Anthropofagist Manifesto* celebrated adage, written originally in English, "Tupy or not Tupy, that is the question" published in 1928. It is not only an allusion to Shakespeare's Hamlet. According to Nodari/Amaral (2018, 2462), '... the most famous aphorism of Oswald de Andrade's *Manifesto Antropófago* raised a question that ended up being obliterated through its conversion into a motto of identity' ("[...] o mais célebre de seus aforismos converteu aquilo que se apresentava como dilema em um lema identitário").

13 *Atas da Câmara Municipal de São Paulo*, 1562–1903.

'Hence, Tapajós, Pauxis, Surubiú, Maturu, Gurupatuba are nowadays called, respectively, Santarém, Óbidos, Alenquer, Porto de Moz, Monte-Alegre [...]'[14]

Nevertheless, what causes the replacements is not a longing for Portugal, as indicated by Antenor Nascentes (1960, 103), but the obliteration of Native cultures. Thus, northern Brazil, where previously the Grão Pará Province had been established, retains Portuguese toponyms results of that 18th-century policy: Santarém, Belém, Barcelos, Chaves, Óbidos, Viseu, among many others—all Portuguese toponyms.

In addition to the Portuguese, other Europeans had contact with the Natives, including the French. They came to Brazil to establish settlements in the north (Equinoctial France, today's Maranhão, 1612–1615) and Rio de Janeiro (Antarctic France), as observed by French chroniclers Jean de Léry (a Calvinist) and André Thevet (a Franciscan).

The French established themselves on an island in the Guanabara Bay in 1555 (today Ilha de Villegagnon). A few men were sent to the mainland to make bricks for the construction of houses and a fortress. These men settled between two hills called Mont-Henri and Mont Corguilleray (currently *Morro da Viúva* 'widow's hill' and *Morro da Glória* 'Glory's hill'). The names represented tributes to the King and the sponsor of the expeditions, Philippe du Pont de Corguilleray (Carvalhinhos/Antunes 2007, 151). Henriville is a toponym that refers to a small and underdeveloped settlement on the coast (in contrast to the island's settlement), according to Knauss (2008, 145).

The names of the island and the old fortress were a homage to Admiral Coligny. The local brickworks had been named *Briqueterie*, a common noun: the flooded island *Ratier* (Fr., 'mousetrap'), and the Sugar Loaf as *Pot de Beurre* (Fr., 'butter pot'), both names referring to the rock shape (Léry 1980, 102, 106 and 249). The only tangible evidence of French place names nowadays is the Ilha de Villegagnon (formerly, Coligny Island) within Guanabara Bay in Rio de Janeiro. Other foreign influences appear later in the 19th century, during the period of transition between Monarchy and Republic.

At the end of the 16th century, northern Brazil had still not been explored and occupied by the Portuguese. This attracted the attention of French explorers. In France, the political and religious upheavals of the time resulted in King Henri VI's murder. Louis XIII assumed the throne. His mother, Maria de Médici, the Regent and devoutly Catholic, supported Mr. de la Ravardière (Daniel de La Torche) and Admiral François de Rasilly's trip to Maranhão. Capuchin priests led by D'Evreux accompanied them on the journey, and their exploratory report resulted in d'Abbeville's text, *Histoire de la Mission des Peres Capucins en l'Isle de Maragnan et terres circonvoisines*. São Luís, the capital's name, bears witness to that time. According to the IBGE, the town was raised to the status of a city on the 9th of August, 1612. Abbeville's report on the French's

14 "Daí, pois, Tapajós, Pauxis, Surubiú, Maturu, Gurupatuba se chamarem hoje, respectivamente, Santarém, Óbidos, Alenquer, Porto de Moz, Monte-Alegre, [...]"

movement describes how Fort São Luís was established as a colonists' defense post. A tribute to Louis XIII, the fort's name, also was transferred to the village.

The toponymic memory of Dutch Brazil (or New Holland, 1630–1654, Mello 2010) is limited to Pernambuco (Recife and Olinda), even though the West Indian Company undertook raids throughout Bahia. Johan Maurits van Nassau-Siegen (Dutch Brazil's Governor, 1637–1644, Mello 2010), known as Maurício de Nassau, features in numerous urbanonyms (Google Maps 2022), these including Uninassau (a college in Olinda), Maurício de Nassau school, colleges, a university center, residential condominiums, a driving school, tourism agencies, etc.

Guararapes (from Tupi *GUARARA, a bird, + [a]pé [r, s], 'guarara's path' according to Navarro 2013, 563). The author explains [r, s] as different prefixes that this word accepts (see Navarro 2013, 563). This a toponym that recalls two decisive wars (Guararapes Wars) against the Dutch, in 1648 and 1649 (Mello 2010). The Guararapes Mountains were the scene of these battles. The name spread extensively, but it is impossible to determine whether in commemoration of the battles or by reference to the hills, i.e., Jaboatão dos Guararapes (municipality), Guararapes Mountains, nowadays part of the Guararapes National Historical Park, Monte dos Guararapes Station, Guararapes College, Memorial Guararapes Cemetery, among many others. The most famous urbanonym is Recife International Guararapes Airport, recently renamed Guararapes-Gilberto Freyre Airport.

Other toponyms from Dutch Brazil are linked to 16th and 17th-century fortresses. *Forte Orange* is the toponym to Forte São João de Itamaracá (Ilha de Itamaracá, PE), and *Orange* was also used for inns and hotels in the area (Orange Praia Hotel, Itamaracá). The name Orange refers to the Dutch royal house Orange-Nassau. Another fortress is Forte São João Batista do Brum, the current name of the former Forte Novo de São Jorge, also Forte Diogo Pais (1629). The word "Brum" present in the current name is a Luso-Brazilian voice and part of the new name after the Dutch invasion in the fortress in 1630. The Dutch changed the fortress name to *Fort de Bruyn* in homage to the Dutch War Council President, but Luso-Brazilians called it *Brum*, and not *Bruyn*. The names of some ruined fortresses also bear witness to Dutch Brazil, such as Castrum Giselini/Gyselingh, Fort Thyszzon (Cabo de Santo Agostinho, PE), and Castro Dussem (Albuquerque/Lucena/Walmsley 1999).

3.1.1 Current toponyms

The concept of toponymic areas connects naturally to two further concepts, dialectal areas and cultural regions. Carvalhinhos (2014a, 1050) defines these as follows:

> 'Concentrations of place names united by ethnic, social, economic, political, or even physical distinctions characterize toponymic areas usually connected with cultural areas. By identifying such areas, toponymists obtain (from the semantic substance of the nomenclature) clues about the

namegiver group, its reality and its way of seeing and decoding the physical and cultural world.'[15]

The dialectal areas of some Native languages correspond to toponymic areas. Dick (1990) used the cultural regions defined by Manuel Diegues Junior to relate them to the Brazilian toponymic areas resulting from economic or historically relevant regions. Her list, however, does not consider the first economic activity of the colony: the extraction of brazilwood. There was no impact or naming of areas based on this, perhaps because the entire territory had received its name derived from that activity. Dick (1990) highlights other place name areas derived from the environment, such as those related to Brazilian flora. We have, however, decided to look more closely at the resulting areas of Brazilian historical and economic cycles:

- *The Northeast coastline*, the first area of colonization. From Sergipe to Rio Grande do Norte, place names of the northeastern coastal area are marked by the economy of sugarcane cultivation and processing. Plantation locations, villages, and constructions related to this economic cycle also appear in toponyms: Vereda da Canavieira 'sugarcane brook', BA, Canavieira 'sugarcane', PI, Ilha Cana Brava 'wild sugarcane island', BA.
- Northeast, hinterland. The livestock industry creates place names like Vaca Morta 'dead cow', a town in PE, Riacho Bezerro 'calf creek', BA, Curral de Cima 'upper cattle shed', PB, Curralinho 'small cattle shed', BA and Boi Morto 'dead ox', a town in PI.
- Midwest region/Minas Gerais. The place name corresponds to the second historical and economic cycle, that of mining (gold and precious stones). Minas Gerais has the highest concentration of such toponyms: Diamantina (something like 'made from diamonds'), Lavras 'mining', Cristais 'crystals', Cristália (-ália is a suffix to build toponyms and means 'land of'), Ouro Branco 'white gold', Ouro Fino 'fine gold', Ouro Preto 'black gold', Prata 'silver' and Pratinha 'silver + diminutive'.
- States of São Paulo, Rio de Janeiro, and Paraná. In the late 19th and early 20th centuries, these states introduce toponyms related to coffee culture. Japanese and Italian settlers came to Brazil to work in the coffee plantations, and they represent an essential human component that currently constitutes a significant part of the population of the state of São Paulo. These place names are related to coffee activities, such as Cafeara, Cafelândia 'coffee land' and Cafezal do Sul 'southern coffee

15 "[...] las áreas toponímicas ligadas a áreas culturales se caracterizan por concentraciones de topónimos unidos por matices étnicos, sociales, económicos, políticos y hasta incluso físicos. Mediante la identificación de tales áreas el toponimista obtiene (a partir de la sustancia semántica de la nomenclatura) pistas sobre el grupo denominador, su realidad y su manera de ver y decodificar el mundo físico y cultural."

plantation', all towns in the State of Paraná; Cafelândia, the name of a city in the state of São Paulo; Rubiácea 'Rubiaceae', SP.
– Southern Region. Toponyms related to cattle and horse raising: Invernada Grande 'big winter pasture', PR, Pasto Ruim 'poor pasture', stream, RS; Touro Passo 'bull's passage', RS, Potreiro Grande 'big paddock', RS.

3.1.2 Current areas of Native place names

Particularly in urban spaces, the existence of a Native toponym does not necessarily indicate any former Native existence in the place.[16] First, we need to establish Native toponymic areas related to current dialectal areas. Usually, anthropologists outline ethnographic regions, which interconnect with dialectal areas. Only after this is it possible to assess whether we might be dealing with an original nomination.

The first area to be highlighted is in the coastal region. Genuine Tupi toponyms remained in the Brazilian coastal landscape, between Maranhão and São Paulo States (at São Vicente). In the manual *Método Moderno de Tupi Antigo*, Navarro ([3]2008, 11) states the following:[17]

> 'Since the early colonization of Brazil, people spoke the same language along the Brazilian coast, from Pará to the south of the country, around the parallel of 27 degrees south (according to information by the chronicler Pero de Magalhães Gândavo). Cardim tells us that it was spoken by the Tupiniquim, Potiguara, Tupinambá, Temiminó, Caeté, Tabajara, Tamoio, Tupinaé and other peoples. In the 16th century (and even more in the 17th century), the Portuguese called this language *língua brasílica*' (author's highlights).

Traditional peoples had named settlements and rivers before the arrival of the Europeans. Coastline river names had been transferred to human geographic features, such as Pernambuco, Paraíba, Sergipe, Itamaracá, the rivers Potiguar and Capibaribe[18] (the latter in Recife), all in northeastern Brazil, among others.

---

16 The policy for naming thoroughfares in Brazil—squares, streets, avenues, squares, roads, and paths—is linked to the proposal of bills by council members, representatives elected by popular vote.

17 "Desde os primeiros tempos de colonização do Brasil, constatou-se que, na costa brasileira, desde o Pará até o sul do país, aproximadamente até o paralelo de 27 graus (segundo informações do cronista Pero de Magalhães Gândavo), falava-se uma mesma língua. Cardim nos diz que ela era falada por tupiniquins, potiguaras, tupinambás, temiminós, caetés, tabajaras, tamoios, tupinaés, etc. Já no século XVI e, mais ainda, no século XVII, foi dado a ela pelos portugueses o designativo de *língua brasílica*."

18 Pernambuco: PARANÃ PUKA 'hole in the sea'; Paraíba: PARÁ AÍB A 'bad river'; Sergipe: SERI ÎY -PE 'at the crab river'; Itamaracá: ITÁ MARAKÁ 'stone rattle'; Potiguar: POTIGÛARA 'shrimps eater'; and Capibaribe: KAPI'IBARA 'Y-PE 'at the grass eater river'.

The hydronyms of the cities of São Paulo and Rio de Janeiro retain Tupi names: Rio Carioca (KARIÎÓ-OKA 'carijó's house', RJ), Rio Maracanã (MARAKANÃ 'parrot', RJ), Rio Tietê (TY ETÉ 'magnificent river', SP), Rio Anhangabaú (ANHANG-Á OBÁ 'Y 'devil's face water', SP), Rio Tatuapé (TATÚ (A)PÉ 'armadillo's path', SP), among many others. Overall, they are descriptive toponyms that mark aspects of the material and spiritual cultures of Native groups. From the coast of São Vicente (SP) to the south of Brazil, place names from Guarani, a branch of the Tupi language, were given by Natives during the colonial period.

In Brazil's hinterland, there is a predominance of ethnic groups whose languages belong to another linguistic branch, the *Macro jê* (formerly called *gê*), languages commonly spoken by the Tapuia. Here are some examples, listing Native toponyms areas whose names had probably been coined by Native peoples (i.e., they are not artificial toponyms):

The Coastline, from Maranhão to São Paulo: The Tupi people. Examples: Baía de Sepitiba (SEPÉ -TYB A 'bay of lots of grass', RJ), Barra do Cunhaú (KUNHÃ 'Y 'women's river mouth', RN), Córrego Taquaruçu (TAKÛARA UÂSU 'big taquara stream', ES), Ilha de Urubuqueçaba (URUBU KER -SAB 'island of the place where vultures sleep', SP), Praia de Itapuã (ITÁ PU' Ã 'beach of the raised stone', BA), Rio Ipojuca ('Y APÓ -ÎUK A 'river water from rotten roots', PE), Rio Tracunhaém (TARACU NHAEN 'ants pan', 'anthill river', PB.)

Northeast (plateaus and valleys of the São Francisco River): Kariri (extinct language). Examples: Rio Bodó (RN), Bodocongó (PB), Rio Moxotó (PE).

The Coastline, from São Paulo to the South: Guarani. Examples: Ivoti (RS), Itatí (RS), Pirapó (PR/RS), among others.

Midwest, languages from the Macro jê linguistic branches, especially Bororo. Examples: Torixoréu ('dark hill', MT), Baá-Curiréu (historic toponym of Cuiabá, the capital, meaning 'big settlement', MT), Cudorórí (hill at Chapada dos Guimarães, from CUDORO 'blue macaw' and RI 'hill', MT).

Amazon: Karib (Rio Aicurú AI-CURÚ 'wild dog river'), Arwak toponyms (the Arwak name to the Amazon River, Guiena, from UÊNI 'river', is quoted several times in the First Treaty of Saint Ildefonso, 1777, according to Cardoso 1961, 389).

To recognize the common noun for water in place names may help researchers to establish Native toponymic areas. Brazilian Native languages have agglutinative structures, making it difficult for Portuguese speakers to identify generic (geographic) and specific (the name that particularizes it) features. Thus, isolating the generic term might be a useful approach, one which is especially helpful in the Amazon and adjacent regions. Since Native languages are semantically opaque to Portuguese speakers, these toponyms generally receive a new Portuguese generic term, concentrating two generics into one toponym. The following is a list of water-related terms from non-Tupi[19] languages:

19 All non-Tupi etymologies come from Cardoso (1961).

Karib toponyms.[20] The words *tuna*, *curu*, and *paru* mean 'water'. Examples: the rivers Arái Paru (from Macuxi ARÁI 'piranha fish' and PARÚ 'river', AM), Tucá-Paru (from Karib TUCÁ 'Brazil nuts' and PARU, AM), Conobô-tunã (from Macuxi CONOBÔ 'rain' and TUNA 'river', AM), and the igarapé Curu (AM) and river Parucotó (from Karib PARÚ + COTÓ 'family', AM).

Arwak toponyms: many words express the concept 'water', such as *ári*, *uiná*, *uene*, *uáu*. Examples: the rivers Saué-uiná (from Parecí SAUÊ 'lots of grass' and UINÁ 'river', MT), Badecure Uáu (from Wapixana BADECURE 'jaguar' and UÁU 'river', AM), Poduáua (from Arwak PODÚ 'black' and ÁUA 'river', AM), Curicuri-Ári (from Arwak CURICURÍ 'parakeet' and ÁRI 'river', AM), Sipuene (from Baniwa SIPÚ 'vulture' and ENE 'river', PA), Anauene (from Arwak ANÁ 'flower' and UENE 'river', AM).

Bororo toponyms: the Bororo people use the words *iáo* and *bo* 'river'; *po* or *poba* 'water'; *baga* 'stream'; *paro* 'river mouth.' Examples: the rivers Côgue-iáo (from CÔGUE 'golden fish' and IÁO 'river', MT), Biriqui-páro (from BIRÍ 'leather' QUI 'dry' PÁRO 'stream', MT), Poxoréo (PÔ 'water' and XORÉU 'dark', MT), Padarôbo (PADÁRO 'froth' and BO 'water', MT), Curugúga-bága (CURUGÚGA 'sparrow-hawk' and BÁGA 'stream', MT), etc.

Brazilian hydrography is a good source to recognize the toponymic constituents from Native languages, many of them from the Tupi or the general language, as clearly shown by the examples mentioned above.

3.1.3 Native toponymic name-forming elements

Linguists (Rodrigues 1986; Melatti 2007) divide the dominant Native toponymic name-forming elements into two groups: one derived from nouns and adjectives, the other derived from other elements, such as adverbs. Generally, it is not easy to identify Tupi adverbs. As mentioned, Brazilian Native languages have an agglutinating structure, different from BP, which is a fusional language. Ordinary speakers often interpret the whole syntagma as a single word. We list the primary name-forming elements, all of them nouns, related to environmental or cultural elements.

3.1.3.1 Water: 'Y, PARÁ, PARANÃ

'Y. Used initially in Tupi for all-natural terms or body fluids, it assumes the meaning of 'liquid'. It was also used to mean 'water' in the general language. *'Y* is not associated directly with a concrete phoneme in Portuguese (↗4 Historical phonetics and phonology). This explains why *i* and *u* alternate in place names according to the phonetic realization in each region. Navarro (2008, 14) explains:

20 *Karib antigo* 'Old Karib', according to Cardoso (1961).

> 'We symbolize with *y* a particular phoneme that does not exist in Portuguese, but which is found in both Russian and Romanian. In phonetic transcriptions, usually, it is represented by ɨ: *ibityra* [iβɨtɨra]—*hill*; 'Y [ʔɨ]—*water* is a mid-vowel, intermediate, between **u** and **i**, with the tongue placed as in **u** and outstretched lips for **i**' (author's highlights).[21]

At the beginning of the word. According to Tupi grammar, *'y* can appear at the beginning of the word when accompanied by an adjective, particularly in descriptive toponyms: Rio Iguaçu ('large water river', PR and AM), Baía do Iguape ('at the river cove bay', BA), Ipiranga 'red water' (town, MA, stream, SP), Córrego Ipanema ('unlucky water creek', MT), and Upanema ('unlucky water creek', RN), Uberaba 'shiny water' (town, MG).

At the end of the word. In these cases, *'y* appears after the noun, expressing the relation between the two words. Rio Andaraí ('bats river', BA), Rio Pacuí ('pacu river', MG—pacu is a colossal fish typical of Amazonian rivers), Sarapuí ('sarapó river', town, SP—Sarapó is a fish from Gymnotidae family), Tatuí ('armadillo river' town, SP).

In the middle of the word. *'Y* also appears in middle position, such as in Rio Capibaribe ('at the capybaras' river', PE—capybara is a large semi-aquatic rodent from South America), Peruíbe ('at the shark river', SP), Itaípe ('at the stone river' BA), or is converted into *gy* in place names such as the rivers Sergipe ('at the crab river', SE) and Cotegipe ('at the agouti river', PR).

Pará, Paraná: These words mean 'large river' and 'river', respectively. The term *paraná* denotes rivers in northern Brazil. It does not appear as a generic term in other regions of the country but rather as a place name. Paraná also assumes contracted forms such as *perná* (Pernambuco, PE) or *parná* (Santana de Parnaíba, SP).

3.1.3.2 Natural elements and landscape

There are many landscape elements that generate place names. To illustrate, we have chosen the most representative ones.

The ground. yby. Ibiúna ('dark ground', SP), Butantã ('hard hard ground', SP), Ibitupã ('Tupã's ground', BA).

Hiding places. kûara. Jabaquara ('fugitive's lair', SP), Araraquara ('macaws burrow', SP), Jericoacoara ('turtle hole', CE), Urubuquara ('buzzard hole', island, PA).

Stones. itá. Itacurussá ('stone cross' island, RJ), Itacoatiara ('painted stones', small town, AM).

Bushes. ka'a. Capão ('bush island', small town, BA), Capão Alto ('elevated bush', SC), Caatinga ('white bush', river, MG).

---

21 "Representaremos com *y* um fonema que não existe no português, mas existe no russo e no romeno. Em transcrições fonéticas, geralmente representa-se por ɨ: ibityra [iβɨtɨra]—montanha; 'Y [ʔɨ] —água. É uma vogal media, intermediária entre *u* e *i*, com a língua na posição *u* para os lábios e os lábios estendidos para *i*."

Wood. ybyrá. Ibirapuera ('old trees', SP), Ibiraúna ('dark wood', BA), Ibirarema ('stinking tree', SP).

Fish. pirá. Pirajubá ('yellow fish', MG), Piracicaba (from pirá sykaba[22] 'place where the fish arrive', SP). Also used for specific species: îundi'a ('catfish river', Jundiaí, SP), îa-u (river, Jaú, SP; îa-u 'gobbler' is a kind of gilded catfish).

Reptiles. mboîa 'snake': Mogi (from mboîa Gy 'snake river', SP), Embu (mboîa 'y 'snake river', SP), Boipeba (from mboîa peb-a 'the flat snake', island, BA); îakaré ('alligator'): Jacareí ('alligator river', SP), Jacareacanga (îakaré akanga 'alligator head river', PA).

Birds. îaku 'guan', Penelopinae: Jacuí ('guan river', BA), îambu ('screamer', Anhimidae family) Anhembi (îambu 'y, SP), Inhambupe (îambu 'y pe 'at the screamer river' river, BA); gûará 'egret' Guarapiranga (gûará pirang-a 'red egret', SP), Araçoiaba (ûara aso'îaba 'egret's feather cloak', SP); arara 'macaw' Ararendá (arara ena 'macaw landing place', CE), Araritaguaba (arara itá 'u-aba 'place where macaws eat saltpeter stones', SP).

Mammals. kapibara 'capybara' (Capivari, kapibara y 'capybara river', SP); îagûara 'jaguar': Jaguaré (îagûara rem 'stinking jaguar', SP), Jaguareguava (îagûara 'y 'ûaba 'place where jaguars drink water', SP); tamandûá 'anteater' Tamanduateí tamandûá eté 'y 'genuine anteaters river', river, SP).

3.1.4 The most frequent grammatical components: adjectives, adverbs, suffixes

Mirim/*-im*, adj.: 'small'. This may occur as a whole word (*mirim*) or as a suffix (*-im*; Itaim 'little stone', from itaĩ). In ancient documents, it appears with a tilde (*ĩ*). Native Tupi elements or other Native languages also have this adjective: Mogi-Mirim 'small snake river', SP; Cururu-Mirim 'little frog', MT; Cuiabá-Mirim kûiaba mirim 'little gourd', river, MT. Portuguese words and toponyms also admit the adjective *mirim*: Café-Mirim 'little coffee', small town, MG; Castanha-Mirim 'little chestnut', lake, AM.

Guasu (also *guaçu*/*-asu*/*-açu*, *-uasu*, *-uaçu*), adj.: 'big'. This may be written either with a cedilla (ç) or with *s*, although standard BP indicates the use of the cedilla, simply because this is older and mostly used in documentation of the colonial period. Examples: Mogi-Guaçu (mbôia [g] 'y -ûasu 'big snake river', SP), Guaçu (creek, MS), Acará-Açu (akará asu 'big scalar fish', small town, PA), Açú (RN), Taquaruçu (takûara/a usu 'big canes', stream, PR).

Qualities. *Porang*, adj.: 'pretty'. Iporanga ('y porang/a 'pretty river', SP), Nuporanga (nhũ porang-a 'pretty field', SP), Ibiporanga (yby porang-a 'beautiful land', BA).

22 This is not a word from the general language, but from the *meridional general language*, a southern general language.

Aýb a 'bad': Rio Parnaíba (PARANÃ AÝB A 'bad river', SP), Paraíba (PARÁ AÝB A 'big bad river', a Brazilian state, and also a river).

Colors. Three colors are the most common in Brazilian toponyms: *ting* 'white', Tabatinga (TOBATING A 'white clay', SP), Capetinga (KAPI'I TING A 'white grass', MG). *Una* 'black' Unaí (UNA 'Y 'black water', MG), Rio Paraibuna (PARÁ AÝB A UNA 'big bad black river', SP). *Pyrang* 'red' Ipiranga ('Y PIRANG A 'red river', SP).

Intensifiers. *-etá*: adj. 'many' (number, countable). In Tupi, the notion of intensification is different from that in Portuguese. Thus, *-etá* and *-etê* suffixes are adjectives (according to Navarro [3]2008), even though in Portuguese the same concept corresponds to another grammatical class, the adverb. Examples: Paquetá (RJ), PAK (A) ETÁ 'many pacas', *paca* is a small nocturne rodent; Caetá (BA) KA'A ETÁ 'many bushes'. *-eté*: adj. 'many' (intensity): Caetité (BA) KA'A ETÉ-ETÉ 'great woods'; Caeté (sierra, MT) KA'A ETÉ 'genuine woods.'

Circumstantials. *-(s)aba*, *-çaba* and others, which indicates place: Guaraqueçaba (GUARÁ KER -SAB 'place where guará birds sleep', PR); *-ava* Igarapava (YGARA UB/A -ABA 'place where canoes berth', harbor, SP); *-ndava*, Avanhandava (ABÁ NHAN -ABA 'place where men run', SP); *-ndaba* Potirendaba (POTIRA -NDABA 'place with flowers', SP); *-aba* Piracicaba (PIRÁ SYKABA[23] 'place where fishes arrive', SP); *-guaba* Tujuguaba (TEÎU 'Y 'U -ABA 'place where lizards drink water', SP); *-guava* Jaguareguava (a peak of Cubatão, São Paulo's State; as explained above, this comes from ÎAGUARA Ý 'ÛABA).

Collective nouns. Collectivity is expressed by the suffix **-***tyba* and the variations *-tuba*/*-tiba*; *-dyva* (*-nduva*/*-ndiva*). According to Navarro ([3]2008, 14), the phoneme [ʔɨ] denotes the collective, meaning 'cluster, many'. Airituba (AÎRY[24] TYBA 'cluster of palm trees', ES), Itatiba (ITÁ TYBA 'cluster of stones', SP), Catanduva (KA'A ATÃ TYBA 'cluster of hard bushes', SP).

3.2 Toponyms of African origin

The arrival of Africans in Brazil as slaves under inhumane living conditions (↗2 The social history of Brazilian Portuguese; ↗7 The history of the lexicon)[25] did not lead to the emergence of place names of African origins, as might have been expected with such a massive influx of people (about four million). Contrary to the case with European settlers, their immigration was forced.

---

23 According to Navarro (2013, 593), SIK was a verb used in the 18th century meaning 'fish arrive', and comes from meridional general language (↗3 The history of linguistic contact underlying Brazilian Portuguese).

24 AÎRI: "wild palm species [*Astrocaryum aculeatissimum* (Shott) Burret], also *brejaúva* or *brejaúba* (V [ocabulário na] L[íngua] B[razilica], II, 63". Italics in the original. Translated from Navarro (2013, 18).

25 The trafficking of human beings to Brazil took place between 1502 and 1860, after which it was banned. The law known as *Lei Áurea* 'Golden Law' legally outlawed slavery in Brazil in 1888.

Lima-Hernandes/Carvalhinhos (2019) bring to light several aspects of BP formation in an interesting discussion of concepts such as *motherhood language* and *heritage language*. In this context, they also study the contribution of the African languages, which is vital in understanding why Brazil has so few African toponyms.

The following factors may explain the low frequency of toponyms of African origin in Brazil's toponymic system: i) the slavery situation, ii) the complex state of African dialects, iii) the subsequent outlawing of the trade, and iv) missing records.

Lima-Hernandes/Carvalhinhos (2019, 127–128) discuss another reason: obviously, the ability to speak local languages and the avoidance of languages considered as foreign led to more opportunities (which might seem awkward given that this was within a situation of slavery). This and all the previous reasons serve to explain the scarcity of toponyms deriving from African languages.

In Bahia, the former presence of Sudanese groups explains African toponyms related to the candomblé religion, mainly in urbanonyms. Across the country, there are repeated loanwords that came from Bantu (probably Kimbundu). These regionalisms appear in place names such as *cacimba* 'well', *mocambo* 'cove, shelter', *quilombo* 'organized shelter', in Kimbundu 'union', Cacimba (PI), Mucambo (CE),[26] Mulungu (CE) and Quilombo (SC).

It is not uncommon to find African toponyms for Brazilian streets. These place names are artificial, however, because they are not the result of the presence of African languages, but rather examples of politically motivated naming. They may be of Bantu or Sudanese origin. Brazilian street names notably make use of deity names, more precisely the names of the following orishas: Oxum, Ogum, Oxumarê, Oxóssi.[27] São Paulo (SP), Nova Iguaçu (RJ), and Vila Velha (ES) all have streets with names from Candomblé.

---

26 Petter (2006/2007, 67–68) notes that the common name for a place of refuge was *mocambo*, Kimbundu, 'den, refuge'. The term *quilombo*, from Kimbundu, also contains the idea of refuge, even though it was essentially a multi-ethnic military institution. Petter quotes Beatrix Heintze, 'an Angolan historian of reference' ("historiadora de referência para Angola") (Petter 2006/2007, 68), who states: 'quilombo (kilombo) [comes] from Kimbundu: Lumbu 'surrounded hedge'. ("quilombo (kilombo): [é] proveniente do kimbundu: lumbu 'cercado, sebe'" (Heintze 1985, 126). And Petter continues: 'The author also states that the term was used by the Angolans from Sao Tomé, possibly of Umbundu ethnicity. They used it to designate villages built on the island around 1544, when they remained following a shipwreck [...] in the Umbundu language, the equivalent term is 'ochilombo' *oci-lombo*, meaning 'night shelter')' ("A autora também informa que o termo quilombo foi utilizado pelos angolares de São Tomé, possivelmente de etnia umbundo, para designar as aldeias que construíam na ilha, por volta de 1544, quando aí permaneceram em razão de um naufrágio. [...] em umbundo, o termo é 'ochilombo', oci-lombo, significando abrigo de noite.").

27 Orishas are gods of natural elements, or animal or ancestral spirits. Oxum ([oˈʃũ] is a water divinity, from the river Ossum in Nigeria), Ogum ([oˈgũ] is the deity of iron users (and also of forges). By extension, it is also the god of war and agriculture, hunting, and of all other activities involving the manipulation of iron tools, Oxumarê ([oʃumaˈre] is a Daomé god identified with rainbows and snakes, representing eternity, movement and continuity, Oxóssi ([oˈʃɔsi] is Ketu's king, the hunter.

4 Post-colonial toponymy

4.1 European immigrants and the Republic

After 1822, post-colonial Brazil changes and there is a need to create a national identity. The Native people become the representative figures of a return to a glorious past, as may clearly be seen in the emergence of Brazilian Romanticism. This movement contributed to the recovery of Native place names that had been given Portuguese forms under Pombal (↗2 The social history of Brazilian Portuguese). The retrieve of original toponyms occurred in the country as a whole. Seraine (1960), for example, provides details on the recovery of Native place names in Ceará.

Another political event that influenced the shaping of Brazilian toponyms was the abolition of slavery, which led to the need to "import" new workforce. Because of this, after 1888, Europeans began to immigrate to work in the fields. The first German immigrants in fact arrived in Brazil in 1824 to work at the Real Feitoria do Linho Cânhamo.[28] After 1850, the immigration process increased, and new settlements emerged in the current southern region. These settlements were given Italian and German toponyms.

Unlike the prescriptive process experienced in northern Brazil in the 17th century, the foreign toponymy of the 19th and 20th centuries is characterized by a longing for the homeland. Thus, many toponyms are marked by the adjective *new*, which is added to the original place name. As examples, we can point the towns Novo Hamburgo, an homage to the German city in the state of Rio Grande do Sul, and Nova Trento, Nova Veneza, Nova Bréscia, and Nova Pádua, all in Rio Grande do Sul, as tributes to Italian cities.

The homages and references to the foreigners in southern Brazil can also be seen in another way, namely through references to the settlers' original regions, such as Westfália (RS) and Pomerode (SC). There are also hybrid compositions, such as São João do Polêsine (RS). A few more examples: Blumenau (SC) and Schroeder (SC), German anthropotoponyms; Brochier (RS), one of the few references to France; and Chiapetta (RS), an Italian deanthroponymic toponym. Sometimes official sources do not present the motivation of toponyms, or present not trustable etymologies. A good example is Witmarsum,[29] a small city in Santa Catarina. Witmarsum's first colonization

28 This Portuguese state-owned company, which owned many slaves who worked in flax and hemp rope production, also produced material for sails and naval rigging.

29 IBGE offers two etymologies without quoting the sources: 'Russian immigrants from the Ukraine arrived six years later and named the place Witmarsum, which means 'blue star'. There is, however, a version based on research by a Dutch university that also explains the name: Witmar would be the name of the founder of the Mennonite religion and 'sum' means garden.' ("Imigrantes russos oriundos da Ucrânia chegaram seis anos depois e deram ao local o nome de Witmarsum, que significa 'estrela azul'. Há, porém, uma versão baseada em pesquisas de uma universidade da Holanda que também explica a denominação: Witmar seria o nome do fundador da religião menonita e 'sum' significa jardim",

took place in 1924. As a municipality, it has existed since 1962. According to IBGE, the former name of Witmarsum was *Nova África* 'new Africa', a name given by the first German settlers. Ukrainians arrived in 1930 and renamed the place.

The place name Witmarsum appears twice in Brazil, since it also names a neighborhood in Paraná. Colônia Witmarsum was built in the town of Palmeira, where Mennonites from Santa Catarina's Witmarsum had bought a farm (Fazenda Cancela 'gate farm'). Nowadays, this neighborhood is known for its cultural heritage (there is a trilingual school) and paleontological peculiarities (glacial striations), two features that increase the locality's potential for cultural tourism. On the other hand, Colônia Witmarsum has a successful dairy industry (brand Cancela), founded by Witmarsum (SC) migrants in the 1950s.

Finally, we must note that after the foundation of the Republic in 1889, the current urban toponymy began to take shape. Whereas it is possible to see Brazil's ethnolinguistic formation by analyzing its natural features, its urban toponymy, especially that relating to hodonymy, reflects new influences and political interests.

4.2 Urban toponymy and political influences

The toponymy of Brazilian cities is exceptionally diverse. In general, following Dauzat's (1926) assumption, the names given to water bodies tend to remain, given the obvious permanence of these geographical entities. Thus, the names given to a town's waterways tend to be rather old. In the case of Native toponyms, they may indicate naming prior to the colonial period, at least in the older settlement areas.

In the transition from the 19th to the 20th centuries (Republic), industrialization and other changes in the modes of wealth production led to the creation of place names. Some villages became cities: for example, São Paulo is currently a metropolis with a population of almost 20,000,000. However, in 1840 it was a small village with an urban area of around 40 streets. The current metropolis integrates small, old urban clusters that in the 19th century were far away from the village itself. Over the course of that century, new neighborhoods filled the vast areas between the original Piratininga village (São Paulo's former name) and these distant clusters. The *sesmarias*' allot-

https://cidades.ibge.gov.br/brasil/sc/witmarsum/historico, last accessed 18.02.2022. However, there is no resemblance to east Slavic words for 'star' or 'blue' in the place name and calling the migrants 'russian' can be misleading since they were primarily speakers of Germanic varieties. The Dutch etymology is more convincing since part of the migration originated in Frisia, but the explanation is reported erroneously. The founder of the Mennonite religion was Menno Simons who happened to be born in Witmarsum when the place name already existed for centuries. According to van Berkel/Samplonius (2018), the place name is documented since the 13th century and the etymology could be paraphrased as 'Witmar's home', '-um' for 'home' or 'residence place' being a common component in Frisian place names (see also Versloot 2011, 130).

ments[30] became farms, giving birth to smaller properties (small farms) and many modern neighborhoods.

As a result, both the urban area and its toponymy grew in an irregular and disorderly way. Somehow, toponymic traces from the old 18th and 19th centuries village remain. The old downtown area (the colonial town's core) also has some place names that refer to this different era of urban life, such as Rua Tabatinguera (TABATINGA -ÚÊR -A 'abandoned clay pit'),[31] Ladeira Porto Geral 'main harbor slope',[32] Rua da Quitanda 'greengrocer street', among others. However, similarly to the nature of its settlement, the city's hodonymy is heterogeneous.

After the Republic settlement (1889), paying tribute to prominent figures becomes the key naming motivation. The city's disorderly expansion also contributed to this. Toponyms often reflect stories of political favoritism and the exchange of favors between those in power. Some downtown public spaces also reference the Republic, such as the street names Quinze de Novembro (date of the Republic's settlement in 1889), Benjamin Constant, Quintino Bocaiúva, among other famous figures of the transition from monarchy to republic.

The dynamics of naming in uptown areas largely reflect the usage of land at the time, based on spontaneous occupations or the allotment of land. Anthroponyms represent a large body of the sixty-five thousand names for public areas here. These include names incorporating specific semantic fields (birds, Native groups, precious stones, etc.) that might indicate both the will of those who name a place and the result of a policy of toponymic structuring of spaces.

In the 1970s, twenty thousand streets either had no name or there was duplicate naming. To solve this problem, the São Paulo City Council created a database for names, a list of artificial toponyms used to generate new names. In most cases, the chosen toponym is not linked to the history and urban development of the neighborhood in question. In this new set of place names, the government also proposed many Native names.

Therefore, it is evident that in these cases, the toponym, despite its Native linguistic affiliation, is neither a result of the immediate existence of Native people nor indicates any sort of historical continuity. There are no toponymy commissions in Brazil, and city council members are responsible for naming, both proposing new place

30 *Sesmaria* is a system created in Portugal during the 14th century to distribute lands for farming. The same system was introduced in Brazil the 16th century and prevailed until the mid-19th century.

31 Tabatinguera was a place near the Tamanduateí River, where people extracted white clay. According to some authors, this was used to paint cottages. Maps of the 19th century indicate that Tabatinguera is close to the river. The street layout remains.

32 Ladeira Porto Geral is a busy, sloping shopping street, descending toward the street 25 de Março, in the old downtown. The latter street was built over the Tamanduateí River, the course of which had been channeled. The name of this sloping street marks the location of what would be the main port of the Tamanduateí River until the mid-19th century.

names and replacing them (through bills and laws). For further information about São Paulo's toponymy, see Carvalhinhos (2011; 2012a; 2012b; 2018), Carvalhinhos/Lima (2013a; 2013b; 2014), Lima/Carvalhinhos (2018), Carvalhinhos/Lima-Hernandes/Lima (2017; 2018).

5 The grammar of toponyms

Brazil's toponyms follow Portuguese rules of grammar. A noun that can be modified usually constitutes the name *par excellence*. A simple name (noun + noun, such as Córrego Atoleiro 'quagmire stream', MT, or Rio Areias 'Sand River', TO) constitutes the simplest structure. In this typical structure, adjective nominalization also occurs.

The second most used structure is noun + adjective, a classic structure of Brazilian toponyms: Arraial Novo ('new village', populated place, MG), Rio Arrepiado ('shivering river', PA), Rio Azul ('blue river', AC, MT, PA, PR, and other States), Rio Bonito ('pretty river', SP, SC, PR), Capela Velha ('old chapel', populated place, RS). A lack of knowledge of Native languages tends to lead to the accommodation of Native toponyms following the syntactic structure of the Portuguese language, resulting in the redundancy of the generic term: Rio Ipiranga ('ʏ river, PIRANG(A) red, 'red river river'). Place names can have inflections or involve a more complex syntactic structure, but there is a tendency to simplify the oral expression. The presence of a noun in the first syntagma may cause the use of an article before the toponym, and this is especially the case in toponyms. Examples here include Brazilian state names: *o* ('the') Rio Grande do Sul, *o* Rio de Janeiro, *a* Bahia, *o* Recife.

The same reasons explain some regional pronunciations of toponyms, such as Salvador, Bahia's capital, a catholic reference to Jesus/Our Savior. Among regional speakers, this vocative also receives a definite article. They call the city *O Salvador*, while other Brazilian speakers generally omit the article. However, *São Paulo*, *Santa Catarina*, and *Belo Horizonte*, place names beginning with adjectives, do not carry articles. This seems to be a rule of toponymic grammar (see also Cunha/Cintra [2]1985, 220).

Occasionally, articles may precede compound nouns, such as *Minas Gerais* and *Espírito Santo* (noun + adj.). There is also a tendency to use articles with Native toponyms which are semantically opaque to most Brazilian speakers: ('the') *o* Piauí, *o* Ceará, *o* Tocantins. However, this is not a rule; for instance, the capital of Mato Grosso, Cuiabá, is not preceded by an article.

Alongside the phenomenon commonly known as onymization, Dick (1990) uses the term *toponimização* 'toponymization', frequently reproduced by Brazilian authors. Toponymization[33] (something like 'the conversion into a toponym') identifies place

33 Although the term *toponimização* is widely used in Brazil, it can be problematic, mainly if the term *false generic* (defined in the geographical nomenclature UNGEGN) is also considered. The phenomen-

names that result from a shift from common nouns to proper nouns. Concerning the use of articles in place names, there is no clear rule from the point of view of toponymic grammar in BP. According to Cunha/Cintra ([2]1985, 220), a definite article must proceed geographic names, but they mention exceptions that reject the article without further explaining the reasons. So, what seems to be a rule of toponymic grammar cannot in fact be considered as such.

5.1 Local Creations

Typical Brazilian creations that involve cultural and ethnic miscegenation have already been described in other chapters of this book (↗1 Brazilian Portuguese linguistics: an overview; ↗2 The social history of Brazilian Portuguese). Whether in the macro space or in cities, each region's place names integrate the virtual naming lexicon of groups, and new toponyms are created from this knowledge. In this sense, Native place names are also used to make new toponyms, although people do not know their etymology. Uncountable hybrid toponyms are generated using morphemes or single words. The most frequent hybrid structures are: i) P + N, ii) N + P, and iii) N + Pm, where P = Portuguese word, N = Native language word (generally, but not exclusively, Tupi), and Pm = Portuguese morphemes or final inflections. There are also hybrid toponyms with African elements (mostly from Kimbundu), although these are only a few that can be found repeatedly.

In what follows we will present examples of typical Brazilian hybrid toponymic creations. The respective place name structure will be given and a few placenames with the same structure will be listed. Whenever the linguistic origin is not Tupi, this will be noted.

5.1.1 Native word first position/BP word second position

Syntagmatic independence: two overlapping independent terms or two words connected by prepositions. Examples: Itapecerica da Serra (SP), from ITÁ 'stone' + PEB 'flat' + SYRYK 'slippery' + A, 'flattened slippery stone' (Navarro 2013, 574) + Port. *da Serra* 'from the hills', a reference to moss-covered stones. Place names with similar structure: Itaguaçu da Bahia (ITÁ ÛASU 'big stone' + da Bahia 'from Bahia 'bay', BA), Itaju

on described by the term *toponimização* presupposes the spontaneous oral formation of a geographical nomenclature which crystallizes through use. However, to prove that a geographical feature incorporating the toponym's specific term is the result of the process of orality is difficult, sometimes impossible. It would lead us to the *false generic*, that UNGEGN (2002, 12) defines thus: "Generic element that does not indicate the feature class of the item named. Examples: Mount Isa, 'Ayn as-Sulṭān, Redhill and Rio de Janeiro are all populated places, not a mountain, spring, hill or river, respectively".

do Colônia (ITÁ ÎUB 'yellow stone' + do Colônia 'from Colônia River', BA), Piraí do Norte (PIRÁ 'Y 'fish river' + do Norte, 'northern fish river', BA), Guaraciaba do Norte (KÛARASY ABA 'Northern sun feather', a kind of hummingbird, CE), Caçapava do Sul (KA'A ASAB ABA 'place where to cross the woods', do Sul 'from the South', 'place to cross the woods in the South', RS).

On the morphemic level: the first term, originating from any Native language, receives a suffix, or a word of Portuguese origin is added. Buritizal (SP), from MERITI, a variant of palm, BURI ('YBA) *-z* (consonant connector) + *-al* (suffix, collection). Other examples: Igarapé Paruzinho, PA (from Karib PARU 'water', 'little water igarapé': paru + Port. diminutive suffix *-inh* + gender inflection *-o*); Rio Parauapebas, PA (PARÁ POPEB A 'large rivers': parauapeba + Port. pl. *-s*.); Córrego Cajueiro, GO (AKAÎU '*Annacardium occidentale* tree and fruit', Navarro 2013, 21; caju + Port. *-eir -o*, suffix used to shape plants' names, among other kinds of constructions in whose it can assume other meanings); Taquaralzinho, MS (TAKÛAR A 'bamboo pipeline', + *-z* consonant connector + *-inh* diminutive + *-o* masculine gender inflection).

5.1.2 BP first position / Native language second position

In all examples of this type, the two parts of the compound toponym are syntagmatically independent. All toponyms present a river name as the second element, except for the four first examples:

Port. + Kaingang: Campo Erê, town, SC. Port. campo 'field' + Kaing. ERE 'beautiful'.

Port. + Kaingang: Nova Erechim, town, SC. Port. Nova 'new' + RE 'field' + sĩ 'small'.

Port. + Tupi: Pedra do Indaiá (MG). Port. pedra 'stone' + de (preposition) + -o (definite article masculine singular) + Tupi INÂIA 'palm tree'.

Port. + Tupi: Cachoeira Bicho Açu (AM). Port. cachoeira 'waterfall' + Port. bicho 'animal' + Tupi UÂSU > açu 'big'.

Port. + Tupi: Cabeceiras do Paraguaçú (BA). Port. cabeceira 'river head' + Port. *de* (preposition) + *-o* (definite article masculine singular) + Tupi PARÁ + UÂSU 'big river'.

Port. + Tupi: Açude São Pedro da Timbaúba (CE). Port. açude 'weir' + São 'Saint' + Pedro 'Peter'+ *de* (preposition) + *-a* (definite article feminine singular) + Tupi timbaúba *TIm̨o'ĩu < TI'm̨o 'timbó', leguminous plant + ĩUA 'plant' (Cunha 1985, 769).

Port. + Tupi: Riachão do Jacuípe (BA). Port. riach-ão (noun + augmentative inflection) 'brook' + Port. *de* (preposition) + *-o* (definite article masculine singular) + Tupi jacuípe ÎAKU 'Y -PE 'in the Penelope river'.

Port. + Tupi: Rio Suaçuí Grande (MG). Port. rio 'river' + Tupi Suaçuí SÛASU'Y 'deer river' + Port. grande 'big'.

5.1.3 Regional hybrid toponyms with African elements

The following examples are among the most productive regionalisms, seen in several Brazilian regions.

With independent words: Mulungu do Morro ('mulungu tree of the hill', BA), Mocambo do Espírito Santo ('Holy Spirit cave', PA), Cacimba do Boi ('ox well', PE).

With morphemic elements of BP: Quilombinho (quilombo + *-inho*, diminutive suffix, MT), Cubatão (KUBATA quimb. 'home' + *-ão* Port. augmentative, SP), Cacimbinha ('little cacimba', BA).

5.2 Popular etymologies and associative meaning in Native place names

Popular etymologies always emerge. Speakers tend to interpret place names according to their own experience. It is common to hear popular etymologies such as 'the Corumbá is over there' for the toponym Corumbataí 'Corumbatá river', interpreted as *Corumba tá aí*, colloquial for *Corumba está aí* 'Corumba is there' or Piracicaba, in popular etymology 'Ou pira, ou se acaba', 'either you become mad or it is over.' The actual etymology for the toponym Piracicaba is 'fish harvest,' a natural place where fishes become stuck, an easy place for fishing.

Popular etymologies tend to create a (false) story that justifies the place name. These imagined etymologies may arise from toponyms of any linguistic origin. Thus, in the perceptions of ordinary speakers, the Native toponym (with a meaning that is usually unknown) is only a *signifier* in the Saussurean sense. Ordinary speakers might know that *corumbatá* is a fish, but they do not know the original meanings in the Native language, or their etyma.

Another case is associative meaning (Nyström 2016, 48). It "implies that a name user – or a group of name users – when hearing a certain name comes to think of something else or something more, apart from reacting to the primary function of the name, namely the function of individualizing and localizing the referent". It makes people understand the toponym in terms of things related to it: Tietê River means, by associative meaning, floods, pollution, stink, heavy traffic. Associative meanings are common and occur in toponyms with different Native linguistic origins.

5.3 Popular regressive derivations

Several issues may lead to regressive derivations. *Sampa*, for example, was popularized by the homonymous song by the famous Brazilian singer Caetano Veloso. However, it is not common to hear a Paulistano, a person from São Paulo, call his city by

this name, which does not extend to the homonymous state. The same happens with *Beagá* (spelling of *BH*), a short and famous toponym for Belo Horizonte.

Other types of abbreviations consider the importance of the city or place in scale. Thus, on a national scale, *Rio* will only be the city of Rio de Janeiro, and at the state level, *Ribeirão* will only be Ribeirão Preto, a significant inland city in SP State, or on a local scale, the same toponym *Ribeirão* indicates the city of Ribeirão Pires, in the Greater São Paulo.

5.4 Suffixes -polis and -lândia, the 20th-century toponymic "boom"

In addition to verifiable toponymic inflection in place names of spontaneous origin (speech origin), it is also necessary to explore some common structures in artificial constructions in Brazilian toponymic grammar. Brazil underwent several changes in the organization of the territory after 1889 (Republic). The division of large tracts of land, the exploitation of territories, the creation of new human settlements on trade routes, the building of railway stations and churches, and the emancipation of rural districts—all this led to successive territorial reorganizations which resulted in toponymic changes. In this sense, several Native toponyms applied to the names of towns and their streets and public spaces are artificial.

The most common toponymic constituents here are the suffixes *-polis* and *-lândia*, vernacular adaptations of the Greek *polis* and the Germanic *land*. These suffixes, among some less common ones (but which are also relevant, all meaning 'land of': *-ália*, *-ino*/*-ina*, *-ínia*, *-ésia*, *-éia*), are almost exclusively found on modern artificial toponyms. The least common constituent (as a suffix) is *burgo*: Cordisburgo (MG), Felisburgo (MG), Luisburgo (MG), Fraiburgo (SC), and Arceburgo (MG). *Burgo* appears in transferred toponyms such as Novo Hamburgo, from Hamburg (Germany), evincing a foreign place name.

Common compositions:

Anthroponym + *-lândia* or *-polis*/*-pole*: Janiópolis (PR), Anitápolis (SC), Andrelândia (MG), Joselândia (MA), Wanderlândia (TO), Augustinópolis (TO), Teofilândia (BA), Quiterianópolis (CE), Medicilândia (PA), Florianópolis (SC), Petrolândia (PE), Carvalhópolis (MG).

Products from the land or plants/trees + *-lândia* or *-polis*/*-pole*: Cafelândia (PR), Marmelópolis (MG), Hidrolândia (GO), Cacaulândia (RO), Buritinópolis (GO), Niquelândia (GO), Babaçulândia (TO), Palminópolis (GO).

Reference to other toponyms + *-lândia* or *-polis*/*-pole*: Acrelândia (AC), Itaipulândia (PR), Tocantinópolis (TO), Israelândia (GO).

Reference to gentilics + *-lândia* or *-polis*/*-pole*: Baianópolis (BA), Candangolândia (DF).

Common words + *-lândia* or *-polis*/*-pole*: Cordislândia (MG), Agricolândia (PI), Curionópolis (PA), Cosmópolis (SP), Divisópolis (MG), Recursolândia (TO), Retirolândia (BA), Rurópolis (PA), Sertanópolis (PR).

Considering only single-word toponyms with such suffixes in the names of 5,537 municipalities (IBGE data, 2011), we can confirm that *-polis* (or *-pole*) appears in 45 percent of them, and *-lândia* in 31 percent. These are the most common toponymic constituents and appear in spontaneous and non-official toponyms such as Cracolândia ('land of crack cocaine', a place in São Paulo city where drug addicts congregate) (see also ↗5 Historical morphology).

5.5 Brand names, acronyms, and other compositions

Brand names and also some artificially-made toponyms are undoubtedly an onomastic expression of social history. Although they are not very frequent in Brazilian toponymy, we can cite some remarkable examples.

Mato Grosso (MT) has been re-explored during the 20th century, engendering some uncommon place name constructions. The names of several cities are acronyms, representing 20th-century state enterprises or colonizing groups. Brazilian Legal Amazon, a program developed by the Federal Government in the 1970s, resulted in new towns, such as Brasnorte, Sinop, Confresa and Colíder, four towns whose names are acronyms: Brasnorte comes from *Empresa Brasil-Norte* 'Brazil's North Company'; Sinop means *Sociedade Imobiliária Noroeste do Paraná* 'Northwest Paraná Real State Society'; Confresa is an acronym from *Colonizadora Frenova Sapeva* 'Frenova Sapeva Colonizer'; and Colíder comes from *Colonizadora Líder* 'Leader Colonizer'. Another artificial toponym is Riversul (PR), an acronym of Ribeirão Vermelho do Sul 'southern red creek'.

On the other hand, a brand name is a vital part of any product, and products are part of everyone's life. At streets, supermarkets, or malls, brand names are created to attract and retain our attention, and to convince us to buy a service or product. Although there are brand names that are created for advertising, others seem to be the very expression of a place name because they reproduce the toponym as their brand.

Various famous Brazilian brand names come from Native languages. Significant here are Ipiranga (oil company) and Itaú Bank. Regionally, there are more brands: Iporanga (drugstores), Ypê (chemistry brand), Iguatemi (malls), Ypióca (alcoholic drinks), Catupiry (cheese), Maguary (juices), Itaipava (beer), Café Iguaçu (coffee), Ipanema (shoes), Maratá (juices, teas, coffee, pepper, among others), Serigy (coconut milk), Piraquê (snacks), and Buriti (rice).

A lot of dairy brands seem to reproduce a toponym: Gravatá, Avaré, Camanducaia, Capixaba, Cariri, Ipanema, Itambé, Embaré, Itacolomy, Jaguaribe, Marajoara, Maranguape, Mantiqueira (also guaraná and eggs), Mucuri, Paracatu and Piracanjuba.

Some famous brands are derived from toponyms: Casas Bahia (stores), Casas Pernambucanas (stores), Porto Seguro (insurance company), Pão de Açúcar (supermarkets), and Lindoya, Prata and São Lourenço, three well-known mineral water brands.

Acronyms or compositions are present in the names of a few state-owned companies (some of them currently under a privatization process) and include the lexical root of *Brazil*, *Bras-*: Petrobrás (oil company), Embratel (telecommunications), Embraer (aircraft construction and defense), Eletrobrás (electricity).

To conclude this section, we will mention three other acronym constructions which are well-known brand names: one is a technology company (Itautec—from *Itaú Technology*), another a popular retail group (Magalu, the brand for *Magazine Luíza*), and finally a bank (Bradesco, an acronym of *Banco Brasileiro de Descontos*).

6 Glossonyms and ethnonyms

In general, symbiosis happens between ethnonyms and glossonyms or ethnonyms and anthroponyms. In Brazilian onomastics, this is also the case. The literature on onomastics classifies the study of ethnonyms and anthroponyms as a shared branch of research, although in fact this is not a simple assumption. Koopman (2016, 251) states that

> Ethnonyms fall under the broader onomastic category of anthroponyms [...]. Ethnonyms refer to the type of group known by a wide variety of different terms: race, nation, population, polity, tribe, clan, kingdom, chiefdom, and 'ethnic group'.

The ICOS list of terminology says the following:

> **Ethnonym**—proper name of an ethnic group (a tribe, a folk, a clan, etc.), or a member of this group, i.e., *Italians*, *Bavarians*, *Croat*, *Frenchman*, *Zulu*. (NOTE: Ethnonyms are not treated as proper names in some languages and by some scholars, i.e., *ingleses* in Spanish. According to some theories, ethnonyms are proper names both in plural and singular, in other theories, ethnonyms in the plural are proper names, in the singular appellatives).

Also, in BP (and in all variants of Portuguese), gentilic adjectives express nationality. Rosa (2020, 26)[34] describes the differences between gentilic adjectives and ethnonyms, noting that gentilic adjectives express the geographical location of a group, whereas ethnonyms refer to a unique collectivity and thus receive proper name status. According to the rule set out by the *Associação Brasileira de Antropologia,* 'Bra-

34 Rosa (2020) notes that there is no consensus among scientists here. She considers the two classic documents from the 1950s, the *Convenção para a grafia dos nomes tribais* by the Brazilian Association of Anthropology, and *A grafia de nomes tribais brasileiros*.

zilian Association of Anthropology', Native ethnonyms are always written in the singular, although in normative grammar the plural may also appear in ethnotoponyms.

In 2010, the IBGE census revealed for the first time data about Native groups. The population of self-declared *Traditional people* was around 890,000, distributed across 305 Native ethnic groups and speaking 274 Native languages. There is a widespread false perception that Brazil is a monolingual country. In fact, there are, apart from the Native languages, several regional varieties of BP, plus a BP variety spoken in Quilombola communities. There are also immigration languages (heritage languages) such as Vêneto and Pomerano (Italian and German dialects preserved in Brazil), as well as smaller groups that retain their original languages, these often spoken in home settings by their respective communities (Hungarians, Russians, Arabs, etc.). New linguistic influences continue to arrive with influxes of immigrating groups of Syrians, Haitians, Bolivians and Venezuelans (↗2 The social history of Brazilian Portuguese).

Whether created by the group itself or by external members, ethnonyms carry their original motivation. It may be a reference to a hydronym or an hodonym, as in Tapirapé, TAPI'IRA (A) PÉ 'tapir's path', a description; as in Tabajara, TOBAÎARA 'enemies', or something mentioned in a depreciatory way (as Cambeba, AKANGA PEB A 'flat head'). Referring to, describing, or bullying a foreign group (see the following section) is a common phenomenon for several cultures and can result in the formation of some anthroponyms.

In the 16th century (Souza 1851), Portuguese chroniclers used to describe the Native groups with the noun *gentio* 'gentile' or 'pagan'. With nominal complements, ethnonyms emerged from this type of naming: *gentio tupinambá*, *gentio potiguar*. The author describes the Brazilian shore from the north (currently Amazonas) to the south (Rio da Prata watershed).

Souza (1851, 55) names every Native group that did not speak the general language *Tapuia*, meaning 'pagan foreigner' in Tupi. The author also refers to Native groups from the shores with similar habits and languages, although they were seldom enemies. Some ethnonyms: Potiguar ('the ones that speak Tupinambá and Caeté language'), Caeté, Tupinaé, Amoipira, Ubirajara, Aimoré, Goitacá, Tamoio, Papaná, Guaianã, Maracá, and Carijó (also referred to as Guarani). The author uses the term *Tupinambá language* as a glossonym, employing it more frequently than *Língua Geral*. Alternatively, he uses geographic features in the genitive: *o gentio do Rio Real* ('royal river pagans', border region between the current Sergipe and Bahia).

Historically, foreign peoples tend to create depreciative names to refer to each other. Varnhagen gives the following examples (Souza 1851, 406, comment 221):

> The Tupi call themselves Tupinambá, or 'remarkable Tupi' [...], and to their looser enemies which had been pushed to the South and the Hinterland, they named them *tupi-ikis* [Tupiniquim] and *tupin-aem* [Tupinaé], which means 'underling Tupi' and 'evil Tupi'[35]. (My translation)

According to Rosa (2020, 30), glossonyms are commonly group names, such as Tukano, Mundurucu, Kaingang, Sateré Mawé, Terena, Xavante, Yanomami, Wapixana, Ticuna, Timbira, Baniwa, etc. Regarding cartographic sources, Souto (2017) and Rocha/Presotto/Cavalheiro (2007) describe not only image representations but also the way that foreign cartographers represent the Native peoples. Through iconography or words, the cartographers' vision oscillates between a romanticized and a negative/depreciative view. For examples, *Canibalis* is a term used by some cartographers. The same dichotomy appears among Sousa narratives and those of other colonial chronicles. They refer to Native men as two groups, the evil and the good: *cruel*, *wild*, *mischievous*, or *domestic*, *well-reasoned*, *of good character*.

A glossonym may also be converted into an ethnonym. For example, since the 16th century, the Amazon general language (Nheengatú or *Nyengatú*, from NHE'ENGA KATU 'good language') has changed its status. Nowadays, ethnic groups that had lost their languages (or whose languages are dead) adopted Nheengatú as their language, using the name as a glossonym and as an ethnonym as well (cf. Lima-Hernandes/Carvalhinhos 2019).

7 Anthroponomastics

The structure of anthroponyms in Brazil is composed of *first name* (simple, compound or triple, the latter being rarer) and *surnames*, in first position the mother's surname, which may be abbreviated or omitted, then the father's surname. The number of surnames may vary.

There are no restrictions as to the choice of first names, the same as in Portugal and other countries. All names are accepted, including foreign names. Names that cause embarrassment to the bearer are forbidden (Law 6.015/73), which does not mean there are no names that are considered "exotic" or "strange."

The choice of a first name embraces numerous issues, but the meaning or semantic content has little weight, so we might consider the name as a pure label, as some philosophers have observed (Mill 1867, in the first place), and this seems to make sense in the Brazilian anthroponymic system.

---

35 "[...] Os Tupis, que a si se chamavam Tupinambás ou Tupis abalisados, foram logo seguidos de outros de sua mesma raça, que se chamavam também a si Tupinambás, e deram aos vencidos, que empurraram para o sul o para o sertão, o nome de Tupiikis e de Tupin-aem, isto é, Tupis laterais e Tupis maus [...]." (Vanhargem in Sousa, 1851, 406)

The melody of the anthroponym is one of the most important reasons for its choice, as well as the religious aspect. Tributes to friends and family members, along with the influence of the media, are common. Therefore, anthroponyms are chosen largely for their sound/melody, or in relation to their meaning in the Native cultures themselves.

We should note that members of the general population can normally recognize a Native anthroponym. However, its meaning is probably not transparent to them. Naming motifs may be extralinguistic: a political issue, an influential artist, a personal homage; all these motivations exist.

In Portuguese, the grammatical gender of a noun has its origins in Latin. Masculine inflection: *-o*, feminine inflection: *-a*. Portuguese does not have a neutral gender like Latin (↗7 The history of the lexicon; ↗13 Morphology). However, the vowel *-i* can serve to avoid gender inflections, and it is used widely in anthroponyms of a Native origin. For further information, see Amaral/Seide (2020), a recent study which provides an excellent overview of the Brazilian anthroponymic system.

7.1 Naming in traditional peoples

As reported by the last census (2010), IBGE created a database for consultation. It shows the most common anthroponyms, as well as frequencies, decades, and geographical distribution, among other features. Using this database is possible to list the most common anthroponyms of Native origin in Brazil. Although they are classified by gender, some are used in both genders.

Masculine anthroponyms. Decades: 1930–1940: Peri. 1940–1950: Guaraci, Iberê, Tabajara, Tibiriçá. 1950–1960: Tabajara, Iberê, Jandir, Moacir, Caubi. 1960–1970: Guarani. 1970–1980: Ubiratã, Taiguara, Raoni. 1980–: Caíque.

Feminine anthroponyms. Decades: 1930–1940: Araci. 1940–1950: Jacira, Guaraciaba, Jaci, Jurema, Iracema. 1950–1960: Guaraci, Irani, Moema, Miracema, Jurandira, Jandira. 1960–1970: Juçara, Jurandira, Jandira. 1970–1980 Juçara, Jandaíra, Jandaia. 1980–1990: Iara, Janaína, Maiara, Moema, Moara, Potira, Bartira. 1990–2000: Maiara, Janaína, Moara, Tainá, Taiane, Iandara/Jandara. 2000–: Tauane.

However, Brazilian Native nations can have other naming standards, even using a "Brazilian" name or a "white" name for civic purposes. In the naming systems of Native groups, the anthroponyms provide invaluable material that help anthropologists understand the groups' family relationships and the transmission of names and their worldview. Historically, the Jesuits and the Portuguese Crown erased Native, "pagan" personal names. Since 2012, a legal provision[36] guarantees traditional peoples the

36 The National Council of the Public Ministry approved the *Joint Resolution* No. 3, 04.19.2012. The Resolution sets rules for providing birth certification in the Civil Registry of Natural People from Native

right to choose their children's names in their own languages, but in reality, they face obstacles with civil registration services to name children if their ethnicity is not commonly known.[37]

7.2 Current tendencies

The choice of a first name appears to be influenced, in all social classes, by more subjective than objective questions, ignoring the matter of semantics. Short and traditional first names (João, Pedro, Joaquim, Maria, Clara, Ana) have become the favorite within higher social classes in recent years.[38] However, citizens with a lower level of education tend to avoid this kind of first name. In their opinion, they are "too simple" and only poor people are called João, José, Maria, etc.

It is more likely to find invented, manufactured, or modified first names among persons with lower socioeconomic status. They are generally generated from a known array of components. These names are sometimes written with many consonants, double letters or other elements that supposedly make them sound foreign, more "sophisticated" and that hence bring luck for the bearer. Although there is general a gap with respect to this issue within onomastic research in Brazil, some research data can be shown. Scottini (2011) studied first names and surnames at private and public schools in São Paulo and made several analyses. He developed a statistical index based upon Fryer/Levitt's (2004) model and called it *Índice socioeconômico de nome*, 'first name socio-economical index'—ISN, and *Índice socioeconômico de sobrenome* 'surnames socio-economical index'—ISS.

Analyzing the ISN, the author concludes that 'higher classes parents tend to adopt common first names to their kids, but lower classes parents use to adopt uncommon names' (Scottini 2011, 26).[39] He also notices that there is not an ethnical issue behind this matter because comparing students of African descendant with other racial groups into the same school (public or private) showed no divergence in first name

groups. This resolution states that the choice of first name is free – except if the nominated person might suffer bullying, according to Law No. 6, 015/73. The Resolution also states that if wanted, the ethnicity should be registered as a surname. See article 2, paragraph 1.

37 For example, a BBC's news article from 2016 tells the story of a Kaingang father and his fight to register his son with a Kaingang anthroponym, Kágfer: https://www.bbc.com/portuguese/geral-36799201, last accessed 18.02.2022.

38 This assertion comes from the research findings of undergraduate students in their final dissertations under my supervision, and is also a subject discussed during congresses. For further information, the reader can search partial results of a research project coordinated by Dr. Juliana Soledade Barbosa Coelho (Federal University of Bahia/State University of Brasilia), *Novo Dicionário de Nomes em Uso no Brasil* (2017–), https://dicionariodenomesdobrasil.com.br, last accessed 14.03.2022.

39 "pais de maior status tendem a preferir nomes mais comuns, enquanto pais de menor status têm mais chance de adotar nomes incomuns para os filhos."

frequency. However, difference did appear, revealing uncommon first names (unique names), when the author compared public versus private. In private schools, students tend to have traditional names (such as Julia or Gustavo), but they tend to have more uncommon names in public schools. He also provides other information that is similar to our empirical analysis here: '[...] the index based on the dichotomy public/private school reveals the economic situation of the family as well as the education level of the parents' (2011, 19).[40]

Furthermore, indexes do not depend on the frequency of first names in society overall. The lower the index, the greater the chance of being made-up or peculiar first name. Several unique first names have the final inflection -on. As examples, the author quotes first names such as *Elivelton*, *Cleberson*, *Gleison*, *Weveron*, *Josimar*, *Wesley*, as male names (2011, 32), and *Carolaine*, *Karolaine*, *Tamires Aparecida*, *Andressa Aparecida*, *Jéssica Aparecida*, *Daiane Aparecida*, *Gleice*, *Leidiane*, *Adriele*, as female names (2011, 33). Other conclusions also are similar to our own empirical analysis:

> 'It is essential to note that 12 of the 20 first names [the author refers to the "unique" and low-frequency first names] present the final inflection *-on*. Besides, 6 of them have at least a foreign letter (K, W, Y). It shows that the first names of public school students tend to sound like foreign names, mainly from English. On the one hand, we can conclude that the most frequent first names of private school students sound Latin (from Portuguese or Italian). On the other hand, the typical lowed-frequency first names sound anglicized' (Scottini 2011, 31–32).[41]

Interpreting the index analysis, the author concludes that if a first name suffers changes in its spell, it becomes "poor":

> 'The impoverishment of the first names happens when foreigner characteristics increase [when more foreign letters are added to the original name.] For instance, Davi [...], David [...], Deyvid [...], Deivid [...], Deividi [...], and Deivide [...]. There is evidence that massive American culture is especially attractive to the poorest Brazilian when the question is the first name choice. (2011, 32) [...] Why is this kind of first name so popular among the poorest people? Maybe because mass culture (cinema, TV, and music), especially that produced in English, has been a successful paradigm among the poorest or low-educated Brazilian people. Case studies about the popularization of celebrity names, such as Princess Diana (written Daiana, Leididiana,) or Hollywood actors like

40 "o índice baseado na dicotomia escola pública/privada capta tanto uma dimensão de renda da família do aluno quanto a dimensão de educação dos pais."

41 "O interessante é que entre os vinte nomes listados, 12 são terminados com o sufixo ON. Além disso, 6 possuem pelo menos alguma letra estrangeira (K, W e Y). Isso revela uma tendência dos nomes tipicamente de alunos de escolas públicas de também soarem estrangeiros. Mas a influência aqui vem, predominantemente, da língua inglesa. Podemos concluir que se por uma [um] lado os nomes que estão na cauda superior da distribuição ISN das escolas privadas soam latinos (derivados do português ou do italiano), os nomes típicos da cauda inferior soam americanizados."

> Michael Douglas, these case studies show that pop culture is very attractive to the poorest Brazilian' (Scottini 2011, 36).[42]

7.3 Made-up first names

It is not uncommon to find anthroponyms constructed from other anthroponyms. Despite legal protection (the 1973 law mentioned above), there exist artificial names such as the "exotic" made-up names studied by Carvalhinhos (2014b). In this study, Carlielssom (13 years old), Andresca (15), and Alextricia (25) all with created, invented, or manufactured names, are asked about the origin of their names. In the first case, Carlielssom says that his name is a "mix of names that my parents invented." On the other hand, Andresca and Alextricia had names created on the basis of the names of their parents, as if they were making a tribute to them:

> 'The explanation for creating names suits the field of psychology better than the field of linguistics, since when a new matrix is created, as in Alextricia (daughter of Alexandre and Patrícia), the etymological meaning is practically nonexistent, even if we retrieve the names from the creation it was based on. Despite the fact that a new name is a montage from the parents' names as if the child brought the two of them together in one name, it is not possible to have an etymological montage, as the etymological meaning is latent but probably bleached in the parents' personal names' (Carvalhinhos 2014b, 99).

The mechanism of name composition usually follows psychological components. Made-up, bestowing names can express homage, self-mirroring (when parents try to mirror themselves into their children's names), or merely something very meaningful to the parents. For Martins (1991, 31–32),

> '[...] Freud was able to indicate the preferential psychological mechanisms in the composition process of personal names. These mechanisms are condensation and displacement. [...] A description of the mechanism is already a breakthrough in knowledge. He explained the formation of personal names through fusions, condensations, and junctions of various terms. Bestowing names that are based on other names is common: JOSEMAR (José + Maria), MARIANA (Maria + Ana), [...] EDIGÊNIO (Edite + Eugênio). The author of this list of condensations notes that names

---

42 "Esse processo de empobrecimento do nome a medida que ela vai aumentando suas características estrangeiras acontece também com outros nomes como, por exemplo, Davi (61,83), David (33,88), Deyvid (31,35), Deivid (19,92), Deividi (12,89) e Deivide (9,76). [...] Há, portanto, evidência de que a cultura de massa norte-americana é especialmente atraente entre os brasileiros mais pobres quando da escolha do primeiro nome. (2011 31–32) [...] Por que os nomes com tais estrangeirismos tem [têm] apego tão grande entre a população brasileira mais pobre? Uma conjectura plausível é que a cultura de massa (cinema, televisão e música), especialmente a produzida em língua inglesa, tem servido de modelo de sucesso para pais brasileiros mais pobres e/ou menos educados. Estudos de caso como a popularização de nomes como os da princesa Diana ou Lady Di (com grafias como: Daiana, Leidiana) e os de atores de Hollywood como Michael Douglas confirmam que a cultura pop é especialmente atraente nas camadas mais pobres da população brasileira."

are composed with love, "as a kind of additional recognition given to the child through a legal resource, which is the civil registration. It almost corresponds to proof of parenthood when the parents are not married"'[43] (Martins 1991, 31–32).

The press (and now other forms of media) often publishes lists of exotic or non-usual first names, these either a combination of parents' first names or even a random composition of letters or syllables (i.e., from a business consulting client's database). Some examples in Pinhoni (2013): Arunaben, Dagolis, Diortagna, Eliximar, Ediolaque, Giodax, Heluede, Jocifrania, Kecimiylla, Lifelete, Mozayra, Nullyany, Ovelarde, Phanttnopy, Quetillane, Romeniggue, Sansgenos, Tetclaymber, Vyttianne, Wanderquisia, Xilziany, Yerkyleydy, Ziueldo.

No studies in onomastic science have yet interpreted such phenomena, which appear to be quite common in northeastern Brazil.[44] In most cases, the many reports of such namings note the exotic aspect of the names, often taking a popular view or even an unpleasant, critical one. However, it is not enough to collect and list first names without analyzing them; we need to consider extralinguistic factors, such as parents' or namegivers' education level, socioeconomic class, and psychological factors, among others.

Rodrigues (2019) analyzed made-up names from the late 19th to 20th centuries, using registration sheets collected from a religious order in Salvador, Bahia as a corpus. One of her hypotheses is that the northeastern made-up names illustrate a kind of linguistic competence in the use of morphological components, which are taken from German anthroponyms.

To sum up, the following list based on Carvalhinhos (2014b) offers some general aspects of Brazilian anthroponymy:

a. Family traditions/homages are common motifs.
b. It is possible to detect fashionable periods of naming in official anthroponym lists, relating to such things as fictional characters, football players, singers, actors/actresses, among others.
c. Religious motivations may occur, but they were more frequent in the 1950s and earlier.

43 "[...] Freud soube indicar os mecanismos psíquicos específicos preferenciais que formam o trabalho de composição de um nome próprio. Esses mecanismos são a condensação e o deslocamento. [...] A descrição do mecanismo já é um avanço no conhecimento. Ele explicaria a formação de nomes próprios através de fusões, condensações, junções de termos diversos. É frequente a denominação de algumas pessoas a partir de outros nomes: JOSEMAR (José + Maria), MARIANA (Maria + Ana), [...] EDIGÊNIO (Edite + Eugênio). O autor desta lista de condensações observa que são nomes inventados com amor, 'como uma espécie de reconhecimento suplementar da criança, através de um recurso legal, que é o registro civil. Corresponde quase a uma perfilhação, quando os pais não são casados [...]'" (Martins 1991, 31–32).

44 There is no study to prove this. The claim is based on my activity as supervisor of undergraduate students, who often research this subject.

d. It is rare to observe names chosen due to their etymological meaning.
e. Exotic names are relatively frequent, and the bearers usually experience bullying.

8 Conclusions

This overview has described several characteristics of Brazilian onomastics. The chapter's design has echoed the development of research in Brazil: some literature on toponomastics exists, but in general onomastics remains like a new world looking for exploration. In light of space limitations, we have explored here only a few aspects. We began by reflecting on how toponyms might have been before the European invasion. Then, moving through the colonial era, we saw how mixed the Brazilian toponymic system is.

Traditional peoples (Native groups) and the enslaved population of African origin and migrants from Europe brought their languages and habits – and names – to this land. We have discussed the subsequent ideological clashes, shedding light on how different worldviews can mark a toponym. Diachronically, we revisited toponymic traces that are hardly visible anymore today, even for Brazilians, for instance, the French presence in Rio de Janeiro (Ilha de Villegagnon) and Maranhão (São Luís, from Louis, the French King).

Considering onomastics as a particular field of linguistics, we have tried to address linguistic and extralinguistic questions. In this sense, toponyms portray Brazil's economic and historical cycles. We have also explained how Native place names indicate dialectological areas. Linguistically, we have offered elements for a Brazilian grammar of toponyms, underlining Tupi and Portuguese morphological constituents.

The chapter has also considered typical toponymic creations: compositions, derivations, such as *-lândia* and *-polis*, which have enjoyed a massive presence in 20th-century toponymy, and other elements present in acronyms and brand names. The symbiosis between ethnonyms and glossonyms was also noted, as well as the premises of the Brazilian anthroponymic system.

In conclusion, this chapter has sought to provide a broad picture of Brazil's toponyms, names and personal names, although it is a picture taken with a long-distance lens. The reader will recognize the main aspects of Brazil's onomastics and is invited to further explore the many open issues.

9 Bibliography

Albuquerque, Marcos/Lucena, Veléda/Walmsley, Doris (1999), *Fortes de Pernambuco: imagens do passado e do presente*, Recife, Graftorre.

Amaral, Eduardo Tadeu Roque/Seide, Márcia Sipavicius (2020), *Nomes próprios de pessoa: introdução à antroponímia brasileira*, São Paulo, Blucher, https://openaccess.blucher.com.br/article-list/9786555500011-453/list#undefined, last accessed 18.02.2022.

Andrade, Oswald (1928), *Manifesto antropófago*, Revista de Antropofagia 1/1, https://digital.bbm.usp.br/view/?45000033273&bbm/7064#page/2/mode/2up, last accessed 18.02.2022.

Ayrosa, Plínio (1938), *Vocabulário na língua brasílica. Manuscrito português-tupi do século XVII*, vol. 20, São Paulo, Departamento de Cultura, https://purl.pt/28813,last accessed 18.02.2022.

Berkel, Gerald van/Samplonius, Kees (2018), *Nederlandse plaatsnamen verklaard*, Rotterdam, Mijnbestseller.nl.

Biblioteca Digital Mundial (2016), *Carta portulana*, https://www.wdl.org/pt/item/18177/, last accessed 18.02.2022.

Cardoso, Armando Levy (1961), *Toponímia Brasílica*, Rio de Janeiro, Biblioteca do Exército.

Carvalhinhos, Patricia (2011), *Toponímia e memória: quando o passado não é utopia*, in: Rosiane Goncalves Braga/Sheila Dias Maciel (edd.), *Memória e Utopia: experiências de linguagem*, vol. 1, Cuiabá, Editora da Universidade Federal de Mato Grosso, 77–114.

Carvalhinhos, Patricia (2012a), *Toponímia paulistana no século XX: Tatuapé. Uma contribuição para o PHPP*, in: Manoel Mourivaldo Santiago Almeida/Maria Célia Lima-Hernandes (edd.), *História do Português Paulista: modelos e análises*, vol. 3, Campinas, UNICAMP/Publicações IEL, 431–457.

Carvalhinhos, Patricia (2012b), *Instituto Previdência, São Paulo: memória(s) e topônimos*, in: Alcione Tereza Corbari (ed.) *Anais do X Encontro do CELSUL: Círculo de Estudos Linguísticos do Sul*, Cascavel, Universidade Estadual do Oeste do Paraná, 1–17.

Carvalhinhos, Patricia (2014a), *Variaciones lexicales en la hidrotoponimia portuguesa: el genérico*, in: Joan Tort i Donada/Montserrat Montagut i Montagut (edd.), *Els noms en la vida quotidiana. Actes del XXIV Congrés Internacional d'ICOS sobre Ciències Onomàstiques*, vol. 1, Barcelona, Generalitat de Catalunya/Departament de Cultura/Direcció General de Política Lingüística, 1047–1064, http://www.gencat.cat/llengua/BTPL/ICOS2011/108.pdf, last accessed 18.02.2022.

Carvalhinhos, Patricia (2014b), *Brief overview of unconventional (or exotic) first names in the Brazilian anthroponymic system*, in: Oliviu Felecan/Diana Felecan (edd.), *Unconventional Anthroponyms: Formation Patterns and Discursive Function*, Newcastle upon Tyne, Cambridge Scholars Publishing, 77–107.

Carvalhinhos, Patricia (2018), *Política y toponimia en Brasil: la moda del guión*, Rivista Italiana di Onomastica 23/2, 228.

Carvalhinhos, Patricia/Antunes, Alessandra Maria (2007), *Toponímia brasileira. Origens históricas*, Cadernos do CNLF 11/3, 141–158, http://www.filologia.org.br/xicnlf/3/Cad%2003%20XICNLF.pdf, last accessed 18.02.2022.

Carvalhinhos, Patricia/Lima, Adriana (2013a), *Denominaciones en el espacio público brasileño: algunas palabras sobre São Paulo*, Rivista Italiana di Onomastica 19/2, 605–632.

Carvalhinhos, Patricia/Lima, Adriana (2013b), *Names in Brazilian Public Spaces: a few words about São Paulo*, in: Oliviu Felecan (ed.), *"Name and Naming": Proceedings of the Second International Conference on Onomastics: Onomastics in Contemporary Public Space*, vol. 1, Cluj-Napoca, Argonaut, 263–276.

Carvalhinhos, Patricia/Lima, Adriana (2014), *Espacios públicos de São Paulo (Brasil): sus denominaciones*, Revista Trama 10/20, 45–52, http://saber.unioeste.br/index.php/trama/article/view/10344/7460, last accessed 18.02.2022.

Carvalhinhos, Patricia/Lima-Hernandes, Maria Célia/Lima, Adriana (2017), *Lo sagrado que se convirtió en profano en la toponimia de São Paulo, Brasil*, in: Oliviu Felecan (ed.), *Name and Naming. Proceedings of the Fourth International Conference on Onomastics: Sacred and Profane in Onomastics*, Cluj-Napoca, Mega/Argonaut, 575–593.

Carvalhinhos, Patricia/Lima-Hernandes, Maria Célia/Lima, Adriana (2018), *The ideological function in names of public spaces in the city of Sao Paulo, Brazil*, Onomastica Uralica 14, 93–110, http://mnytud.arts.unideb.hu/onomural/kotetek/ou14/08.pdf, last accessed 18.02.2022.

Carvalho, Joaquim Barradas de (1965), *As fontes de Duarte Pacheco Pereira no "Esmeraldo de Situ Orbis"*, Revista de História 31/64, 329.

Couto de Magalhães, José Vieira (1897), *Sétima Conferência para o Tricentenário do Padre Anchieta. Assumpto: Anchieta, as raças e línguas indígenas*, São Paulo, Typographia a Vapor Carlos Gerke & Cia, http://www.etnolinguistica.org/biblio:magalhaes-1897-anchieta, last accessed 18.02.2022.

Cunha, Celso Ferreira da/Cintra, Luís Filipe Lindley ([2]1985), *Nova gramática do português contemporâneo*, Rio de Janeiro, Nova Fronteira.

D'Abbeville, Claude (1614), *Histoire de la Mission des Peres Capucins en l'Isle de Maragnan et terres circonvoisines*, https://gallica.bnf.fr/ark:/12148/bpt6k57399d/f40.item, last accessed 18.02.2022.

Dauzat, Albert (1926), *Les noms de lieux: origine et évolution*, Paris, Delagrave.

Dick, Maria Vicentina de Paula do Amaral (1990), *A motivação toponímica e a realidade brasileira*, São Paulo, Arquivo Público do Estado de São Paulo.

Dick, Maria Vicentina de Paula do Amaral ([3]1992), *Toponímia e Antroponímia no Brasil. Coletânea de estudos,* São Paulo, Serviços de Arte Gráfica da Faculdade de Filosofia, Letras e Ciências Humanas da Universidade de São Paulo.

Dick, Maria Vicentina de Paula do Amaral (2000), *Contribuição do Léxico Indígena e Africano ao Português do Brasil*, in: *VI Congresso da Associação Internacional de Lusitanistas*, Rio de Janeiro, http://www.ebah.com.br/content/ABAAAAnPYAH/contribuicao-lexico-indigena-africano-ao-portugues-brasil, last accessed 18.02.2022.

Diégues Júnior, Manuel (1960), *Regiões culturais do Brasil*, Rio de Janeiro, Centro Brasileiro de Pesquisas Educacionais.

Fryer, Roland/Levitt, Steven D. (2004), *The Causes and Consequences of Distinctively Black Names*, Quarterly Journal of Economics 119 (3), 767–805, https://scholar.harvard.edu/files/fryer/files/the_causes_and_consequences_of_distinctively_black_names.pdf, last accessed 18.02.2022.

Google Maps (2021), *Recife e Olinda*, https://www.google.com/maps/search/Nassau/@-8.0851599, -34.9504903, 11.18z, last accessed 18.02.2022.

Guidon, Niéde (1992), *As ocupações pré-históricas no Brasil (excetuando a Amazônia)*, in: Manuela Carneiro da Cunha (ed.), *História dos índios no Brasil*, São Paulo, Companhia das Letras, 37–52.

Heintze, Beatriz (1985), *Fontes para a história de Angola do século XVII*, Stuttgart, Steiner.

ICOS (2021) = The International Council of Onomastic Sciences (2021), *List of Key Onomastic Terms*, Onomastic Terminology 4, https://icosweb.net/wp/wp-content/uploads/2019/05/ICOS-Terms-en.pdf, last accessed 18.02.2022.

Instituto Brasileiro de Geografia e Estatística (2011), *Índice de Nomes Geográficos*, vol. 1, https://biblioteca.ibge.gov.br/visualizacao/livros/liv56282.pdf, last accessed 18.02.2022.

Instituto Brasileiro de Geografia e Estatística (2015), *Perfil dos municípios brasileiros*, https://biblioteca.ibge.gov.br/visualizacao/livros/liv95942.pdf, last accessed 28.02.2022.

Knauss, Paulo (2008), *No rascunho do Novo Mundo: os espaços e os personagens da França Antártica*, História (São Paulo) 27, 143–153, https://www.researchgate.net/publication/250990671_No_rascunho_do_Novo_Mundo_os_espacos_e_os_personagens_da_Franca_Antartica, last accessed 18.02.2022.

Koopman, Adrian (2016), *Ethnonyms*, in: Carole Hough (ed.), *The Oxford Handbook of Names and Naming,* Oxford, Oxford University Press, 251–262.
Kunstmann, Friedrich (1859), *Atlas zur Entdeckungsgeschichte Amerikas*, Berlin, Asher & Cie.
Léry, Jean de (1980), *Viagem à terra do Brasil*, Belo Horizonte/São Paulo, Livraria Itatiaia/Editora da Universidade de São Paulo.
Lima, Adriana/Carvalhinhos, Patricia (2018), *Toponymic Inflation: When the Politics Dilates Names. The Bridges of São Paulo (São Paulo, Brazil)*, Onomastica Uralica 12, 405–414, http://mnytud.arts.unideb.hu/onomural/kotetek/ou12/28.pdf, last accessed 18.02.2022.
Lima-Hernandes, Maria Célia/Carvalhinhos, Patricia (2019), *O materno, o de casa e o herdado nas dinâmicas espaço-temporais*, in: Patricia Carvalhinhos/Maria Célia Lima-Hernandes (edd.), *A casa, o sapo e o baú. Português como língua de herança*, Rio de Janeiro, Dialogarts, http://www.dialogarts.uerj.br/arquivos/colecoes_ailp/a_casa_o_sapo_e_o_bau.pdf, last accessed 18.02.2022.
Martins, Francisco (1991), *O nome próprio. Da gênese do eu ao reconhecimento do outro*, Brasília, Universidade de Brasília.
Matos, Jorge Semedo de (1999), *A navegação atlântica portuguesa*, Oceanos 39, 82–99.
Melatti, Julio Cezar (2007), *Índios do Brasil*, São Paulo, Editora da Universidade de São Paulo.
Mello, Evaldo Cabral de (2010), *O Brasil Holandês (1630–1654)*, São Paulo, Penguin/Companhia das Letras.
Mill, John Stuart (1867), *A System of Logic, Ratiocinative and Inductive*, New York, Harper & Brothers, http://www.gutenberg.org/ebooks/27942, last accessed 18.02.2022.
Monteiro, John Manuel (2009), *Negros da terra. Índios e bandeirantes nas origens de São Paulo*, São Paulo, Companhia das Letras.
Murault, Malou von (2006), *A árvore que se tornou país*, Revista USP 71, 176–198, https://doi.org/10.11606/issn.2316-9036.v0i71p171-198.
Nascentes, Antenor (1960), *A saudade portuguesa na toponímia brasileira*, Revista Letras 11, http://dx.doi.org/10.5380/rel.v11i0.19903.
Nash, David/Simpson, Jane (2012), *Toponymy: recording and analyzing place-names in a language area*, in: Nicholas Thieberger (ed.), *The Oxford Handbook of Linguistic Fieldwork*, Oxford, Oxford University Press, 392–404.
Navarro, Eduardo de Almeida ([3]2008), *Método Moderno de Tupi Antigo. A língua do Brasil dos primeiros séculos*, São Paulo, Global.
Navarro, Eduardo de Almeida (2013), *Dicionário de Tupi Antigo*, São Paulo, Global.
Nodari, Alexandre/Amaral, Maria Carolina de Almeida (2018), *A questão (indígena) do Manifesto Antropófago*, Revista Direito e Práxis 9/4, 2461–2502, https://doi.org/10.1590/2179-8966/2018/37974.
Nyström, Staffan (2016), *Names and Meaning*, in: Carole Hough (ed.), *The Oxford Handbook of Names and Naming*, Oxford, Oxford University Press, 39–51.
Pereira, Duarte Pacheco (15?? [1892]), *Esmeraldo de Situ Orbis*, Lisboa, Imprensa Nacional, http://www2.senado.leg.br/bdsf/handle/id/242845, last accessed 18.02.2022.
Peters, Elizabeth (2014), *What's in a Name: Toponyms as Linguistic Data for Historical Comparison*, https://www.academia.edu/10410309/Whats_in_a_Name_Toponyms_as_Linguistic_Data_for_Historical_Comparison, last accessed 18.02.2022.
Petter, Margarida (2006/2007), *Línguas africanas no Brasil*, África: Revista do Centro de Estudos Africanos 27–28, 63–89.
Pinhoni, Marina (21.10.2013), *Os nomes mais raros (e exclusivos) do Brasil*, Exame, https://exame.com/brasil/os-nomes-mais-raros-e-exclusivos-do-brasil/, last accessed 18.02.2022.
Posth, Cosimo, et al. (2018), *Reconstructing the Deep Population History of Central and South America*, Cell 175/5, 1185–1197, https://doi.org/10.1016/j.cell.2018.10.027.

República Federativa Brasil (1973), *Lei Nº 6.015, de 31 de dezembro de 1973*, Diário Oficial da União, http://www.planalto.gov.br/ccivil_03/LEIS/L6015consolidado.htm, last accessed 18.02.2022.

Rocha, Yuri T./Presotto, Andrea/Cavalheiro, Felisberto (2007), *The representation of Caesalpinia echinata (Brazilwood) in Sixteenth-and-Seventeenth-Century Maps*, Anais da Academia Brasileira de Ciências 79/4, 751–765, https://dx.doi.org/10.1590/S0001-37652007000400014.

Rodrigues, Aryon Dall'Igna (1986), *Línguas brasileiras. Para o conhecimento das línguas indígenas*, São Paulo, Loyola.

Rodrigues, Letícia Santos (2019), *Neologia antroponímica: o que os nomes de origem germânica têm a nos dizer?,* Master's thesis, São Paulo, Universidade de São Paulo, https://doi.org/10.11606/D.8.2019.tde-09092019-175043.

Rosa, Maria Carlota (2020), *Revisitando a Convenção e A grafia de nomes tribais brasileiros*, Confluência 59, 25–46, https://dialnet.unirioja.es/descarga/articulo/7647915.pdf, last accessed 18.02.2022.

Sampaio, Theodoro (1901), *O Tupi na Geographia Nacional. Memoria lida no Instituto Historico e Geographico de São Paulo*, São Paulo, Typographia da Casa Eclectica, http://www.etnolinguistica.org/biblio:sampaio-1901-tupi, last accessed 18.02.2022.

Scottini, Lucas Costa (2011), *O que o Nome nos ensina? Padrões sociais e raciais de nomes e sobrenomes e performance escolar em São Paulo*, Master's thesis, São Paulo, Universidade de São Paulo, https://www.doi.org/10.11606/D.12.2011.tde-30112011-192644.

Seraine, Florival A. (1960), *Topônimos de Portugal no Ceará*, Revista Letras, Revista da Faculdade de Filosofia da Universidade Federal do Paraná, Curitiba 11, 36–58, http://dx.doi.org/10.5380/rel.v11i0.19901.

Souto, Alanna (2017), *Os indígenas na cartografia da América lusitana*, Boletim do Museu Paraense Emílio Goeldi, Ciências Humanas 12/3, 817–837.

Souza, Camila Fernandes de (2011), *O Esmeraldo de Situ Orbis de Duarte Pacheco Pereira: mudanças e permanências epistemológicas em Portugal no século XVI,* Master's thesis, Campinas, Universidade Estadual de Campinas, https://www.doi.org/10.47749/T/UNICAMP.2011.795500.

Souza, Gabriel Soares de (1851), *Tratado descriptivo do Brasil em 1587*, Rio de Janeiro, Typographia Universal de Lammert, http://www.etnolinguistica.org/local–files/biblio:souza-1851-tratado/souza_1851_tratado.pdf, last accessed 18.02.2022.

UNGEGN (2002) = United Nations Group of Experts on Geographical Names (2002), *Glossary of Terms for the Standardization of Geographical Names*, New York, United Nations Publication, https://unstats.un.org/unsd/geoinfo/ungegn/docs/pubs/, last accessed 18.02.2022.

Versloot, Arjen (2011), *Frisian place-names and place-names in Friesland*, Onoma 46, 127–150, https://doi.org/10.2143/ONO.46.0.2975532.

José Luiz Fiorin

19 Linguistic policy and the orthographic agreement

Abstract: The goal of this chapter is to present the explicit or implicit linguistic policies that have governed the spread of Portuguese in Brazil and informed teaching the language in this country. We discuss the issues of relations between Portuguese and Native languages, including general languages, as well as between the Brazilian and European varieties of Portuguese, during the following main periods: the Jesuit catechism, the Pombaline linguistic policy, the discussion between separatists and loyalists after the Independence, the loyalist victory, the issue of the name of the language spoken in Brazil, the problem of language in the constitutions, the principles described in the National Curricular Parameters, the reaction of the media to what was considered teaching "bad" Portuguese. We present the purist stance that characterized the relations between Portuguese and hegemonic languages: first, French, then English. The orthographic issue is debated. Portuguese had two official orthographies: one from Brazil and one from Portugal and other Portuguese-language countries. We demonstrate how this situation came to be and the political reach of the orthographic agreement, which unified the orthographies, to establish the foundations for creating the Community of Portuguese Language Countries (CPLP). We discuss the reasons the CPLP did not become a truly significant organization and the effective role that Brazil can have in promoting Portuguese.

Keywords: Lusitanization, Brazilianization, mixed languages, purism, prescribed norm, practiced norm, centrality, diversity

1 Linguistic policies and the spread of Portuguese in Brazil

The history of establishing Portuguese in Brazil is long and complex. When the Portuguese berthed here, more than 1,220 languages were spoken in the Brazilian territory (Rodrigues 1993, 90). Until the beginning of the 18th century, no effort was made to spread Portuguese in Brazil. This language was spoken by the elite in certain places on the coast. During this period, the Portuguese crown delegated to the Jesuits the work of catechizing the native peoples and subjecting them to colonial power. The Jesuits, who arrived in Brazil with Tomé de Souza, headed by Manuel da Nóbrega, following the recommendations of the Council of Trent, did their missionary work using the languages of the people they were catechizing. The linguistic policy of the Jesuits was to teach the catechism in the local language:

https://doi.org/10.1515/9783110405958-020

> 'If I do not understand the language of the gentiles, nor do the gentiles understand mine, how can they be converted and brought to Christ? This is why we have the rule and norm of learning the language or languages of the land, where we will preach; and this is the greatest difficulty and the greatest work of this spiritual conquest [...]' (Part IV of the Sermon for Epiphany, Vieira 1959, 23).[1]

Due to this linguistic policy, the Jesuits made grammars and dictionaries of the most commonly spoken languages. For example, José de Anchieta wrote a grammar of Tupi, the most commonly spoken language on the Brazilian coast, published in Coimbra, in 1595 ([7]1990). The most widely spread languages were transformed by the colonizers into general languages.

General languages are the result of colonial intervention, which disorganized the social structures of the autochthonous populations, relocating them and gathering them into villages. This created new socio-interactional relations, which affected the languages involved, causing disappearances, *alloglossia*[2] and bilingualism. A general language was not an original native language, but its evolution based on the speech of *alloglots* and bilinguals (Faraco 2016, 124).

Brazil had two important general languages: São Paulo and Amazonian. By the end of the 17th century, the São Paulo general language was more common than Portuguese in communication among inhabitants of the colony, as shown in the numerous reports from this time:

> 1. 'Certainly, Portuguese families and Indians in São Paulo are so connected to one another today that women and children are raised as mixed in the household, and the language that is spoken in these families is that of the Indians, and the children will learn Portuguese in school' (Padre Antônio Vieira 1694, apud Holanda [3]1956, 174).[3]

> 2. '[...] most of these People cannot speak in another language (the general language), mainly the female and all of the servants, and this inability results in irreparable loss, as we see today in São Paulo with the new vicar who came to that church, and needs someone to interpret for him' (report by Governor Artur de Sá e Meneses 1693, apud Holanda [3]1956, 175).[4]

---

1 "Se eu não entendo a língua do gentio, nem o gentio entende a minha, como hei de converter e trazer a Cristo? Por isso temos por regra e instituto aprender todos a língua ou línguas da terra, onde imos pregar; e esta é a maior dificuldade e o maior trabalho daquela espiritual conquista [...]."

2 Term composed of the Greek *allos*, meaning 'other', and *glossa*, meaning 'language', and refers to a community (i.e., *alloglots*), usually a minority of the population, that uses a language other than the official language of that region.

3 "É certo que as famílias de portugueses e índios em São Paulo estão tão ligadas hoje umas às outras que as mulheres e os filhos que se criam mestiça e domesticamente, e a língua que se fala nas ditas famílias é a dos índios, e a portuguesa a vão os meninos aprender à escola."

4 "[...] a maior parte daquela Gente se não se explica em outro idioma (a língua geral), e principalmente o sexo feminino e todos os servos, e desta falta se experimenta irreparável perda, como hoje se vê em São Paulo com o novo vigário que veio provido naquela igreja, o qual há mister quem o interprete."

Sérgio Buarque de Holanda, based on a series of documents, shows that 'the process of effective integration of the São Paulo people into the Portuguese-speaking world can be said to have happened, in all probabilities, during the first half of the 18th century' ([3]1956, 183–184).[5]

The Amazonian general language was used until the mid-19th century and survives to this day, spoken by small populations in the Rio Negro valley, with the name *nheengatu*.

In the mid-18th century, identifying the use of general languages as the means for missionaries to control the autochthonous population, Minister Pombal, in his confrontation with the Jesuits, published the so-called Directory of Indians (May 3rd of 1757), confirmed by Charter of Dom José I on August 17th of 1758. In paragraph 6, the Pombaline linguistic policy is described:

> 'It was always an unchangeably practiced principle in all nations that had conquered new domains, to thus introduce to the conquered peoples their own language, being indisputable that this is one of the most effective means for rooting out of the primitive peoples the barbarity of their old customs; and having shown the experience, that, whereas we introduce to them the use of the language of the prince, who has conquered them, we also instill in them affection, worship, and obedience to the same prince. Thus, with all the civilized nations of the world observing this prudent and solid system, in this conquest the opposite was done, in which the first conquerors took to establishing in it the use of the language, which they called general; a truly abominable and diabolical invention, such that, in depriving the Indians of all those means, which could civilize them, they remained in primitive and barbaric submission, which has continued until now. To root out this highly pernicious abuse, it will be one of the main responsibilities of the directors to establish in their respective settlements the use of the Portuguese language, not allowing by any means, that boys, and girls, who belong to the schools, and all those Indians, who are capable of instruction in this subject, to use the language specific to their nations, or another called general; but only the Portuguese form, which his majesty has recommended on repeated orders, which until now have not been observed, with full spiritual and temporal ruin of the state' (apud Almeida 1997, 3–4).[6]

5 "o processo de integração efetiva da gente paulista no mundo da língua portuguesa pode dizer-se que ocorreu, com todas as probabilidades, durante a primeira metade do século XVIII."

6 "Sempre foi máxima inalteravelmente praticada em todas as nações, que conquistaram novos Domínios, introduzir logo nos Povos conquistados seu próprio idioma, por ser indisputável, que esse é um dos meios mais eficazes para desterrar dos Povos rústicos a barbaridade de seus antigos costumes; e ter mostrado a experiência, que, ao mesmo passo que se introduz neles o uso da Língua do príncipe, que os conquistou, se lhes radica também o afeto, a veneração, e a obediência ao mesmo Príncipe. Observando, pois, todas as Nações polidas do Mundo este prudente, e sólido sistema, nesta Conquista se praticou tanto pelo contrário, que só cuidaram os primeiros Conquistadores estabelecer nela o uso da Língua, que chamaram geral; invenção verdadeiramente abominável e diabólica, para que privados os Índios de todos aqueles meios, que os podiam civilizar, permanecessem na rústica, e bárbara sujeição, em que até agora se conservaram. Para desterrar este perniciosíssimo abuso, será um dos principais cuidados dos Diretores estabelecer nas suas respectivas Povoações o uso da Língua Portuguesa, não consentindo por modo algum, que os Meninos, e Meninas, que pertencerem às Escolas, e todos aqueles Índios, que forem capazes de instrução nesta matéria, usem da Língua própria das suas Nações, outra

Pombal forced the use of the Portuguese language and prohibited the use of the languages of the autochthonous peoples and of general languages. The first reason listed by Pombal is a civilizing reason, since, according to the Directory, introducing the language of the conqueror is the means to eliminate the barbarity of the old customs of those colonized. The underlying idea is that the language of the stronger is superior to that of the dominated. Secondly, speaking the language of the prince, who conquered them, leads to having affection for him, to worshiping and obeying him. The Pombaline linguistic policy operates on the principle *cuius regio, eius lingua*.

The success or failure of the Pombaline policy is a controversial subject among specialists. Until the 18th century, multilingualism predominated. In addition to the native peoples and general languages, there was a very large population of enslaved Africans who spoke their native languages, as well as lingua francas. In this century, the 'first major wave of Lusitanization of the Brazilian territory' (Lucchesi 2015, 87)[7] took place. However, this spread of Portuguese did not happen because of the 'superiority of Portuguese'[8], a position expressed, for example, by Serafim da Silva Neto (1977, 61), or because of the Directory, but due to economic and social changes. We must consider that there were many difficulties and ambiguities in applying Pombal's determinations: for example, schools were not created and wherever they were established, there were no teachers.

Until the end of the 17th century, the different regions of Brazil did not communicate much with one another. This isolation started to break down with the discovery of gold in Minas Gerais in the 1690s. Large population contingents went there from different places in the colony, as well as from the metropolis (Faraco 2016, 139); trade networks were established to bring supplies to the mine regions, with São Paulo and the South, Bahia and the Northeast, in addition to a connection with Rio de Janeiro, the main port for shipping gold and the new capital of Brazil; an urban society was established on a level that had not yet existed in the colonial society (Faraco 2016, 139–141). These factors led to a hegemony of Portuguese in the Brazilian territory.

If the Directory did not have a central role in spreading Portuguese in Brazil, it did, however, have a significant effect on the discursive construction of the linguistic identity of the country. It silenced the languages of the native peoples, the general languages, the African languages and the country started to consider itself a monolingual country.

---

chamada Geral; mas unicamente da Portuguesa forma, que sua Majestade tem recomendado em repetidas Ordens, que até agora não se observaram com total ruína Espiritual, e Temporal do Estado."

7 "primeira grande onda de lusitanização do território brasileiro."

8 "superioridade do português."

2 Linguistic policy and the construction of a national identity

After the Independence, the name and identity of the language spoken in Brazil began to be discussed. There were two positions on this issue: that of the separatists, who differentiated between the language spoken in Brazil and the language spoken in Portugal, and that of the loyalists, who defended the development of Lusitanity (Dias 2001, 187–188).

The Brazilian national identity started to be constructed during Romanticism. The novel *O guarani*, 'The Guarani', by José de Alencar, for example, creates a myth of the origin of the Brazilian nation. Peri and Cecília form the first couple, a Native man who had accepted Christian values ([19]1995, 268–279) and a Portuguese woman who had accepted the values of nature of the New World ([19]1995, 279–280). Therefore, this nation would be Luso-Tupi, culturally speaking. This foundational myth of Brazil concerns the joining of nature and culture, that is, of American and European values. Thus, Brazil would be considered as a combination of the old and the new world, built after the destruction of the colonial building and the perverse elements of nature. Lusitanian elements remain, though modified by values that are American in nature. The Brazilian nation appears after a flood, the description of which combines the myths of the two civilizations that constitute the Brazilian people, that of Noah and Tamandaré ([19]1995, 291–296).

What is interesting, however, is the role of language in building the national identity. During the process of establishing the European nations, theories were no longer made about the problem of the universality of language, as an element of the constitution of man, endowed with dignity and rights, but the distinctiveness of each national language. Herder (1987, [1]1772; 1996) claims that the soul of a nation resides in the genius of language. Geographical separations of people imply that different communities are influenced by various material conditions, for example, climate and lifestyles, which entails linguistic variations and, consequently, the emergence of different languages, which are national languages. Each language, according to Herder, is a living, organic expression of the spirit of the people. It is the way to know the culture and values of a nation, since it embodies them. In order to establish a nation, according to German philosophy, there needs to be a common language. The Nation-states have a clearly identified national language, standardized by dictionaries and grammars, the teaching of which is one of the foundations of national education.

Therefore, in order for Brazil to consider itself a nation, it had to demonstrate its linguistic identity and, consequently, the distinction between the Portuguese spoken in Portugal and Brazilian Portuguese. It had to differentiate the language spoken in Brazil. Portuguese was spoken, but it was a different Portuguese.

In addition to the foundation of a nationality, *O guarani* shows the identity of the language spoken in Brazil. It is associated to that of the Brazilian man, whose origin is

described in the novel. It is not about the way Portuguese is spoken in Portugal, but Portuguese modified by Brazilian nature (Alencar [19]1995, 116–117). The language spoken in the new country is a reflection, in its pronunciation, syntax and lexicon, of the smoothness and roughness of the nature of America. It is also a fusion of culture with nature. Alencar states that, in addition to the vocabulary, the language "mechanism" is also modified. The novelist asks: 'And how could it be any other way, when an American finds himself in the heart of a rich and opulent nature, subject to new impressions still not translated into another language, in the face of magnificence for which there is still no human verb?' (Alencar 1965, vol. 3, 260).[9] After claiming that Brazil will improve the language, he claims that 'all of the peoples with musical genius have a sonorous and abundant language. Brazil is under these conditions: the national influence has already been perceived in the much smoother pronunciation of our dialect' (1965, vol. 3, 261).[10]

Alencar did not advocate for speaking Tupi, but this modified Portuguese in Brazil. With this concept of people and the language of Brazil, he could never accept that Brazilian literature reproduced Portuguese linguistic canons. It should incorporate the linguistic variety that was spoken in the now-independent country. The linguistic independence of these Portuguese patterns was as important as political independence. This proposal is at the foundation of the long tradition of discussions on the status of the national language, which permeated the entire 19th century, continuing until modern times.

Contrary to this attempt to define the linguistic differences of Portuguese in Brazil in relation to Portugal, the Portuguese and Brazilians in favor of Lusitanization exercised strong pressure, mainly through critiques of Brazilian literary works, in order for the language used in Brazil to remain strictly loyal to the Lusitanian standard. The virulent critiques of the work by José de Alencar serve as an example.

Pinheiro Chagas, a Portuguese writer and critic, in *Novos ensaios críticos*, 'New critical essays', after having complimented the literary value of *Iracema* and the strength of its style, indicated a fundamental flaw in the book, the inaccuracy of the language and a concern with distinguishing "Brazilian" from "Portuguese".

> '[...] the flaw that I see in this tale, the flaw that I see in all Brazilian books, and against which I will not cease to intrepidly protest, is the inaccuracy of the Portuguese language, or at least the habit of making Brazilian a different language from Old Portuguese, through rash and unjustified neologisms, and grammatical insubordinations, which (beware!) will border on the laughable if

9 "E como podia ser de outra forma, quando o americano se acha no seio de uma natureza rica e opulenta, sujeito a impressões novas ainda não traduzidas em outra língua, em face das magnificências para as quais não há ainda verbo humano?"

10 "todos os povos de gênio musical possuem uma língua sonora e abundante. O Brasil está nestas condições: a influência nacional já se fez sentir na pronúncia muito mais suave de nosso dialeto."

> they wish to take on the proportions of an insurrection as a rule against the tyranny of Lobato' (apud Melo ^{3}1972, 11–12).[11]

Henriques Leal, a Brazilian who was extremely zealous of the purity of the language, censors the disregard of Alencar for 'good language'[12] and stands firmly against the thesis of the linguistic distinction between Brazil and Portugal (Melo ^{3}1972, 12–15). In *Questões do dia*, 'Issues of the day', José Feliciano de Castilho, a Portuguese writer, and Franklin Távora, a Brazilian writer, harshly criticized the "mistakes" made by Alencar in his works (Melo ^{3}1972, 15–23). Alencar caused controversy with his critics, mainly in the *Pós-escrito à 2ª edição de Iracema*, 'Postscript to the 2nd edition of Iracema', 1965, vol. 3, 255–266; in the *Pós-escrito à 2ª edição de Diva*, 'Postscript to the 2nd edition of Diva', 1965, vol. 1, 399–406, and in the *Prefácio de Sonhos d'Ouro*, 'Preface of Golden Dreams', 1965, vol. 1, 491–498.

These reactions to the Brazilianization of Portuguese were victorious and the linguistic standard taught in schools and considered correct was European Portuguese. In 1897, Joaquim Nabuco, in his inaugural speech in the Brazilian Academy of Letters, stated:

> 'The Portuguese race, however, as a pure race, is more resistant and thus better preserves its language; we should aim for the uniformity of the written language. We should oppose a messy deformation that is fast among us; we should acknowledge that they are the masters of the sources, which impoverish ours more quickly and we must renew them by going to them. The language is an instrument of ideas that can and should have a relative fixity. On this point, we should strive to support the effort and monitor the works of those who are devoted in Portugal to the purity of our language, in order to preserve the genuine, characteristic, categorical forms of their great era [...] In this sense, there will never be a day in which Herculano or Garrett and their successors no longer have full Brazilian allegiance' (apud Pinto 1978, 197–198).[13]

11 "[...] o defeito que eu vejo nessa lenda, o defeito que eu vejo em todos os livros brasileiros, e contra o qual não cessarei de bradar intrepidamente, é a falta de correção da linguagem portuguesa, ou antes a mania de tornar o brasileiro uma língua diferente do velho português, por meio de neologismos arrojados e injustificáveis, e de insubordinações gramaticais, que (tenham cautela!) chegarão a ser risíveis se quiserem tomar as proporções de uma insurreição em regra contra a tirania de Lobato."

12 "boa linguagem."

13 "A raça portuguesa, entretanto, como raça pura, tem maior resistência e guarda assim melhor seu idioma; para essa uniformidade da língua escrita devemos tender. Devemos opor um embaraço à deformação que é mais rápida entre nós; devemos reconhecer que eles são os donos das fontes, que as nossas empobrecem mais depressa e que é preciso renová-las indo a eles. A língua é um instrumento de ideias que pode e deve ter uma fixidez relativa. Nesse ponto tudo devemos empenhar para secundar o esforço e acompanhar os trabalhos dos que se consagrarem em Portugal à pureza do nosso idioma, a conservar as formas genuínas, características, lapidárias, da sua grande época [...] Nesse sentido nunca virá o dia em que Herculano ou Garrett e os seus sucessores deixem de ter toda a vassalagem brasileira."

This allegiance to Lusitanian standards created a situation in which the standard spoken norm had nothing to do with the standard written norm. Mário de Andrade, in chapter IX of *Macunaíma*, entitled "Carta pras icamiabas", 'Letter to the icamiabas', with a parody of the pre-modernist style, mocks the fact that Brazilians speak in one language and write in another, because Lusitanian standards, considered superior, had the prestige of accuracy (1965, 93–107). The poet ridicules the Lusitanian standards in the parody, showing that, when the São Paulo inhabitants write, they use a language very close to Virgilian, which is called the 'Camões language'[14]. To this end, he uses the second person plural form of address, systematically using the majestic plural, a precious lexicon, even archaic in flavor, employing classical forms of syntax and so forth.

3 The official language and its name

The problem of naming the national language emerged right after the independence. The issue is about creating an exteriority, demarcating a border. We speak a mixed Portuguese, as Alencar had thought, or we speak another language (to see how this issue was addressed in linguistic research, ↗8 The debate on Brazilian and European Portuguese). Portuguese in Brazil was being characterized as different from European Portuguese.

After the Independence, Deputy José Clemente Pereira proposed to the Council that medical surgeon diplomas should be written 'in the Brazilian language, which is the most appropriate' (Dias 2001, 187).[15] A law on October 15th of 1827 established that 'teachers will teach reading, writing [...] the grammar of the National Language'.[16] The issue posed by José Clemente Pereira was not resolved, given the controversy it provoked. Therefore, a circular designation was adopted: the language of the Brazilian nation is the national language (Dias 2001, 187).

While the Portuguese language was established as the official language of Brazil in the Constitution of 1823, the topic did not appear in the Constitution of 1824, granted by Emperor Pedro I. The Republican Constitution of 1891 also makes no reference to the subject. The Constitution of 1934, in item d of the single paragraph of article 150, established that teaching, in private establishments, will be in the national language, except for teaching foreign languages. In the Constitution of 1937, known as *Polaca* 'Polish', which institutionalized the New State, there is no mention whatsoever of the subject. However, in 1938, a strong campaign for the nationalization of education began (Payer 2001, 245–247). Vargas claims that a country is a unit of national

14 "língua de Camões."
15 "em linguagem brasileira, que é mais própria."
16 "os professores ensinarão a ler, escrever [...] a gramática da Língua Nacional."

race, language and thought. Therefore, this national construction needed to start in the school:

> 'Among the measures to take effect immediately, the most relevant refers to the work of nationalization, started in the schools, in some regions where the influx of foreign colonization could create, over time, centers unaccustomed to the rhythm of Brazilian life, due to the persistence of customs, habits, traditions and ways of life peculiar to other races. The language is a noble instrument of national sovereignty. Its spread, in the aforementioned higher-density groups, will develop generations of good Brazilians, in childhood and in adolescence, which, until then, had learned from the textbook of their elders and did not know another history except for that of their ancestors on the other side of the ocean or from other latitudes' (apud Payer 2001, 246–247).[17]

We are far from the period in which "racial" mixing was encouraged by the government and there was a tolerance for using immigrant languages. Brazilianness is an absolute value, which does not encompass diversity under any circumstances; it is a monolithic and centralized unit.

During the second world war, it was formally forbidden for schools to teach in a foreign language, as well as for newspapers to circulate in a non-vernacular language and for church activities and those of other institutions to use languages from other nations. This aimed to target mainly German, Italian and Japanese communities, after Brazil entered the war on the side of the allies. The same had already happened in the first world war, when, in 1917, Brazil declared war against Germany. During the second war, it was also forbidden to speak in public in a foreign language. The most serious thing is that this exclusionary mentality contaminated Brazilians and many immigrants were attacked when they spoke a non-vernacular language.

After the fall of the New State and consequent redemocratization of the country, the Constitution of 1946 transformed the language issue into a constitutional subject. Article 35 of the Act of Transitory Constitutional Dispositions states: 'The Government will name the committee of teachers, writers, and journalists who will express their opinion on the denomination of the national language'.[18]

This committee, composed of five members nominated by the Brazilian Academy of Letters; four from the Brazilian Academy of Philology; the General Inspector of Mili-

17 "Entre as medidas de efeito mediato, a mais relevante refere-se à obra de nacionalização, iniciada nas escolas, em algumas regiões onde o afluxo da colonização estrangeira poderia criar, no curso do tempo, centros estranhos às pulsações da vida brasileira, pela persistência de costumes, hábitos, tradições e modos de ser peculiares a outras raças. A língua é um nobre instrumento da soberania nacional. A sua difusão, nos grupos de maior densidade que acabo de mencionar, formará gerações de bons brasileiros, na infância e na adolescência, que, até agora, aprendiam pela cartilha de seus maiores e não conheciam outra história senão a dos seus antepassados do lado oposto do oceano ou de outras latitudes."

18 "O Governo nomeará comissão de professores, escritores, jornalistas, que opine sobre a denominação do idioma nacional."

tary Education; two University Rectors; the President of the Brazilian Association of Press and two Deputies, approved and sent to the Federal Government a decision written by philologist Souza da Silveira (1960, 291–293), which says that the language spoken in Brazil should be called Portuguese. This position is founded on historical arguments, such as, for example, when Brazil was discovered by Portugal and its language was an instrument of superior civilization, it was imposed on the Brazilian territory. It is also based on linguistic arguments, such as, for example, using a monolingual dictionary when you do not know the meaning of a word, and the differences between the two varieties are minimal, limited to pronunciation and the lexicon. In this decision, the differences are ignored, reducing them to unimportant elements, in order to strongly assert unity.

The conclusion is the following: 'The name of the national language of Brazil is Portuguese Language. This denomination, in addition to corresponding to the truth of the facts, has the advantage of evoking, in two words, *Portuguese Language*, the history of our origin and the fundamental basis of our formation as a civilized people'.[19]

Moreover, there are two other references to the issue of language in the constitution. Article 132 establishes who cannot be registered as voters. Clause II of this article states: 'Those who cannot express themselves in the national language'.[20] Clause I of article 168 stipulates: 'primary education is obligatory and will only be taught in the national language'.[21]

The Constitution of 1967, a charter enacted during the military regime established in 1964, makes two references to the national language. In clause I of § 3 of article 168, it states: 'Primary education will only be taught in the national language'[22]; § 3 of art. 142 repeats the ban on registering 'those who cannot express themselves in the national language'.[23] The issue of naming the national language is no longer presented. It is established that Brazilians speak Portuguese.

It is worth noting that the distinctive denomination of the national language does not always mean respect for the specific ways European languages are spoken in the different countries that had a colonial past. In France, *traduit du brésilien, traduit de l'argentin*, and *écrit en québécois* are used. However, this means of referring to linguistic denominations is linked to the idea that the French spoken in France is pure and true. The rest is nothing more than variation.

19 "O nome do idioma nacional do Brasil é Língua Portuguesa. Essa denominação, além de corresponder à verdade dos fatos, tem a vantagem de lembrar, em duas palavras—*Língua Portuguesa*—, a história de nossa origem e a base fundamental de nossa formação de povo civilizado."

20 "os que não saibam exprimir-se em língua nacional."

21 "o ensino primário é obrigatório e só será dado na língua nacional."

22 "O ensino primário somente será ministrado na língua nacional."

23 "os que não saibam exprimir-se em língua nacional."

In the Constitution of 1988, 'On Nationality'[24] appears in Chapter III of Title II (On the Fundamental Rights and Guarantees), article 13, which states: 'The Portuguese language is the official language of the Federative Republic of Brazil'.[25] This is an absolute value. However, for the first time, the existence of languages of Native American peoples in the Brazilian linguistic reality is acknowledged. Chapter VIII of Title VIII (On the Social Order) is called 'On the Indians'.[26] In it, article 231 recognizes the right of the Native American peoples to preserve 'their social organization, customs, languages, beliefs and traditions [...]'.[27] § 2 of art. 210 establishes: 'Regular basic education will be taught in the Portuguese language, ensuring that indigenous communities can also use their maternal languages and their own learning processes'.[28] This Constitution represents an advancement, since a universal "indigenous community" is not mentioned, but rather 'indigenous communities',[29] thus acknowledging their diversity. The languages of the Native American peoples are recognized as existing in the national enunciative space. However, by asserting the right to language under a title that is not about fundamental rights and guarantees, it establishes a distinction between Portuguese, the national language, and other languages that are acceptable, in a space that is, in fact, seen as monolingual. To separately assert diversity means, in a way, their exclusion. The languages of the autochthonous peoples do not have the same status value as the language of Brazilians in general, since, in the definition of nationality, what counts is Portuguese. Furthermore, the constitution determined that basic education would be taught in Portuguese, which could also be done in a Native American language. The concern is that of integrating all autochthonous peoples in the enunciative space of Portuguese. On the other hand, there has been a radical silencing of immigrant languages and cross-border languages.

Despite the fact that, in Brazil, there are around 180 Native American languages, 2 dozen immigrant languages, LIBRAS (Brazilian Sign Language) and near-bilingualism on the borders, Brazil is seen as a monolingual country. Clearly, today Portugal has been forgotten. There is a Brazilian standard norm that ignores the European standard norm. The correct language is the so-called (Brazilian) standard norm. Nevertheless, there is a difference between the prescribed standard norm and the one effectively practiced: for example, schools teach that you cannot use the forms *ele(s)* 'he/they' and *ela(s)* 'she/they' in the direct object position, though they are currently used in the language of people who use the standard norm (Faraco 2016, 175–176).

24 "Da Nacionalidade."

25 "A língua portuguesa é o idioma oficial da República Federativa do Brasil."

26 "Dos Índios."

27 "sua organização social, costumes, línguas, crenças e tradições [...]."

28 "O ensino fundamental regular será ministrado em língua portuguesa, assegurada às comunidades indígenas também a utilização de suas línguas maternas e processos próprios de aprendizagem."

29 "comunidades indígenas."

What is the linguistic policy in education today? The Ministry of Education presented, in 1997, the *Parâmetros Curriculares Nacionais,* 'National Curricular Parameters'—PCNs, which are guidelines with the goal of '*guiding educators* by standardizing some fundamental factors concerning each discipline' (portal.mec.gov.br/seb/arquivos/pdf/livro01.pdf, last accessed 18.02.2022).[30] Though they are not obligatory, the PCNs are meant to guide educational activities. They are principles for teaching a given discipline, with respect to goals, subjects, ways to implement activities, learning expectations and the types of assessment.

The PCN for the Portuguese language establishes the teaching of reading and writing as the central goal. Regarding speaking, it states that 'it is not about teaching one how to talk or speak 'correctly', but about utterances that are adequate for the context of use' (MEC 1997, 20).[31] Students need to acquire the skill of effective use, in other words, that which achieves the intended effect. Therefore, teaching is based on the assumption that the text is a unit. Since every text is organized in a genre, teaching genres is at the foundation of teaching the Portuguese language. 'Therefore, the school provides the student with access to the universe of texts that circulate socially, in order to teach them how to produce and interpret them' (MEC 1997, 26).[32]

When discussing the use of linguistic varieties, the PCN for the Portuguese language states that people are identified by the way they talk and it is common for less prestigious linguistic varieties to be considered wrong or inferior. The school should discourage linguistic prejudice, since it is one of their goals to 'respect differences'.[33] The goal of teaching the Portuguese language is not to teach students to speak correctly or incorrectly, but to use the variety more appropriately in a communicative situation. 'Speaking well means speaking appropriately; it means producing the intended effect' (MEC 1997, 26).[34]

The school should lead students to use language in different communicative situations, especially more formal ones (MEC 1997, 22). As we can see, the principles for teaching the Portuguese language abandoned the emphasis on correction and proposed focusing on appropriateness. However, they still show that the school aims to teach the use of more formal language in communicative situations in which it is required. However, they insist that dialectal differences should be respected, whether regional or social.

While the media accepts regional varieties of Portuguese and uses them sparingly, it reacts strongly to any attempt by regulatory organizations of education to de-

30 "orientar os *educadores* por meio da normatização de alguns fatores fundamentais concernentes a cada disciplina".

31 "não se trata de ensinar a falar ou a fala 'correta', mas sim as falas adequadas ao contexto de uso".

32 "Cabe, portanto, à escola viabilizar o acesso do aluno ao universo dos textos que circulam socialmente, ensinar a produzi-los e a interpretá-los."

33 "respeito pela diferença."

34 "Falar bem é falar adequadamente, é produzir o efeito pretendido."

fend the respect for colloquial varieties of Portuguese, as was the case in the 2011 controversy regarding the book *Por uma vida melhor,* 'For a better life', by Heloísa Ramos (2009). The book was part of the National Textbook Program, by the Ministry of Education, and was accused by several journalists of leading teachers to teach "incorrect Portuguese" to their students, in place of what they considered "good and correct" language use. These journalists accused the Ministry of Education of allowing the teaching of "bad" Portuguese, in order to keep the population ignorant and, thus, easier to manipulate politically. Despite the media's position, the PCNs remain in effect in schools and most teachers follow its guidelines.

4 The relation between Portuguese and hegemonic languages

The lexicon of any language is always open to loanwords. Portuguese is no exception. It was formed by loanwords from different languages, mainly those that, during a determined period, were spoken by peoples who had political hegemony. In the history of Brazilian Portuguese, these languages were French, at first, and English, later on.

However, at different times, purist movements emerged, which, according to their supporters, aimed to preserve the language from corruption. Foreign linguistic elements are seen as detrimental, a desecration of the spirit of the national language. They promoted activities aimed at eliminating these alien elements. These are times when strong prejudices emerge against everything that is foreign.

When you take, for example, the grammar of Eduardo Carlos Pereira, a book that greatly influenced the education of many generations of Brazilians, in its 102 editions from 1908 to 1950, the following barbarism is found among bad language habits: 'Unnecessary use of foreign terms and terms with a meaning extraneous to the language' ([110]1958, 272).[35] At this time, French was a more influential language on the international stage. Therefore, the grammarian says that 'more than any other language, French is competing to bastardize or barbarize ours' ([110]1958, 272).[36] Though he does not consider himself a purist ([110]1958, 273), the statements above leave no doubt about his position against what he considers to be the bastardization of the language. He presents a list of lexical and syntactic Gallicisms, which, according to him, constitute an example of 'true distortions of the language, against which we should be protected' ([110]1958, 273).[37] For example, according to the grammarian, the following lexical Gallicisms should be avoided: *abat-jour* for *quebra-luz* 'lampshade'; *bouquet* for *ramilhete*

35 "Uso desnecessário de termos estrangeiros e de termos em acepção estranha à língua."

36 "mais do que qualquer outra língua, tem o francês concorrido para abastardar ou barbarizar a nossa."

37 "verdadeiras deturpações da língua, contra as quais devemos estar premunidos."

or *ramalhete* 'bouquet'; *coalizão* for *coligação* 'coalition'; *bizarro* for *esquisito* 'strange'; *desolado* for *aflito* 'distraught'; *debutar* for *estrear-se* 'inaugurate'; *sucesso* for *vitória* 'success'. The following are cases of syntactic Gallicisms: *barco a vela* for *barco de vela* 'sailboat'; *emprestar de* for *tomar emprestado* 'borrow'; *aluga-se quartos* for *alugam-se quartos* 'rooms for rent' ([110]1958, 273–274).

In 1999, deputy Aldo Rebelo presented to the House of Representatives bill nº 1676 to defend, protect, and promote the use of the language. It aimed to establish a linguistic policy and plan, with the lexicon as the area of intervention. It established the requirement to use Portuguese in public places, with the consequent prohibition of using foreign terms.

According to the author of the legal proposition, the language is being invaded by many foreign words, mainly from North American English, which causes two problems: on one hand, a difficulty for our 'simple rural men'[38] to communicate, unaccustomed to the use of foreign terms; on the other, it distorts the language. In the Portuguese/English conflict, a language crisis had emerged. The political and ideological foundation of this policy is to defend the national sovereignty, in the anti-imperialist struggle. Based on this policy, a linguistic plan was established. It determined the requirement, under penalty of a fine, to use the Portuguese language in public domains, except for some cases provided by law. Given the strong criticisms by linguists, the project was not approved in the terms in which it was presented. It shows, however, the presence of purism, which intends to combat linguistic corruption represented by the incorporation of foreign terms.

The purist struggle is in vain, because the lexicon of a language cannot be molded by laws, but is forged by use.

5 The issue of orthographic unification

The orthographic agreement signed by the countries with Portuguese as an official language has been in effect since January 1st of 2009. This convention was approved by Legislative Decree 54, on April 18th of 1995. The duration of the new orthographic norms was determined by Decree 6.583, on September 29th of 2008. The agreement aims to unify the two official orthographies of Portuguese, that of Brazil and that of Portugal, the latter being adopted by the other Lusophone countries. Orthography has always divided Brazil and Portugal.

The different orthographies are based on two principles: a) etymological, which takes into consideration word origin and the way it was spelled in the language of origin; b) phonological, which aims to graphically represent the phonemes that compose

38 "homens simples do campo."

a word. The modern orthography of Portuguese is basically phonological, 'but with remnants from its etymological memory' (Faraco 2016, 191).[39]

Language historians usually divide the history of Portuguese orthography into three periods (Williams 1975, 33–41).

The first is called the phonetic period, which starts with the appearance of the first texts written in Portuguese in the 12th century until the 15th century. During this time, the intent was to spell words as they were pronounced. However, there was no systematic and coherent way to write the words, because there cannot be a phonetic orthography, given the variation in pronunciation. For example, the letter *h* can mark the open vowel or the tonicity of the monosyllable (*he*, *hi*, *hir*); it can take into consideration the etymology of the word (*homẽes*); it can indicate hiatus (*cahir*, *sahir*); it can represent the oral vowel *i* or the nasal *ĩ* (*sabha* = sabia; *camho* = camĩo). Other times, there is no explanation for the use of *h* (*hordenar*, *hobra*). During this period, the same word was spelled in a wide variety of ways.

The second is called the pseudo-etymological period. It lasted from the end of the 15th century until 1911. With the Renaissance, there was a return to Classical Greek and Latin authors, who became objects of imitation. In orthography, in this spirit, there was an attempt to spell Portuguese words similarly to those in Latin and Greek. This is when specific symbols emerged to spell words of Hellenic origin (for example, *pharmacia*, *theatro*, *rheumatismo*, *chimica*, *martyr*). Intervocalic consonants were duplicated (for example, *approximar*, *abbade*), which had been reduced to single ones in the evolution of the language. This period is called pseudo-etymological because many spellings were incorrect, contradicting etymology itself (for example, *chrystal*, *author*, *lyrio*). During this historic period, orthography took on great importance. Numerous orthographers emerged with the goal of systematizing the conventions for spelling words: for example, Pero de Magalhães de Gândavo, Duarte Nunes Leão, Álvaro Ferreira de Vera, João Franco Barreto, Madureira Feijó, and Luís de Monte Carmelo. As Saussure observes, 'the very principle of etymological writing is wrong' (1969, 38),[40] because it disrespects the evolution of language.

The third period, entitled historical-scientific or simplified, started in 1911, when the Portuguese orthography was established. The studies carried out since 1868 by Adolfo Coelho were predecessors of the orthographic changes made during this period. In 1885, Gonçalves Viana and Guilherme Abreu, in the work *Bases da ortografia portuguesa,* 'Foundations of Portuguese orthography', established the principles for an orthographic reform:

39 "mas com resíduos da memória etimológica."

40 "é o próprio princípio da escrita etimológica que está errado."

a) the complete elimination of all symbols that represent Greek phonemes: *th*, *ph*, *ch* [= *k*], *rh* and *y*;
b) a reduction of double to single consonants, except for *rr* and *ss*, which have specific phonetic values;
c) the elimination of null consonants that do not influence the pronunciation of the preceding vowel;
d) a standardization of diacritics (1885).

In 1904, Gonçalves Viana published the work *Ortografia nacional,* 'National orthography', in which a new orthography for Portuguese was proposed. In 1911, the Portuguese government nominated a committee to propose the foundations for the orthographic reform. Among them were José Leite de Vasconcelos, Carolina Michaelis de Vasconcelos, Adolfo Coelho, Epifânio Dias, Júlio Moreira, and José Joaquim Nunes. This committee proposed the adoption of the foundations proposed by Gonçalves Viana, with few changes. This orthography was made official on September 1st of 1911. To complete this reform, Portugal did not consult Brazil. Nonetheless, this orthography was adopted in Brazil in 1931 through an agreement between the Brazilian Academy of Letters, which is responsible for administrating orthography, and the Academy of Sciences of Lisbon. However, the Constitution of 1934, in article 26 of the Transitional Provisions, states: 'This Constitution, written in the same orthography of 1891, which was adopted in the Country, will be enacted by the Board of the Assembly, after it is signed by the Deputies present, and will go into effect on the date of its publication'.[41] Hence, it annulled the agreement of 1931. In 1943, the Luso-Brazilian Convention was signed, which reestablished the agreement of 1931. The Brazilian Academy of Letters approved, on August 12th of 1943, the Orthographic Form, which contains instructions for elaborating the Orthographic Vocabulary of the Portuguese Language. In 1945, due to divergences in interpretation of some orthographic rules, delegates from the two Academies met in Lisbon, from July to October. The result was a document entitled 'Complementary conclusions of the Agreement of 1931'.[42] There were so many modifications that, in fact, it was a new orthographic reform. Since this attempt at unification basically took into account the Portuguese pronunciation, it led to impassioned protests in Brazil. The National Congress did not approve the agreement of 1945. Law 2623, on October 21st of 1955, reestablished the orthographic system of 1943, while, in Portugal, the convention went into effect on January 1st of 1946. It characterized once again the orthographic division.

41 "Esta Constituição, escrita na mesma ortografia da de 1891 e que fica adotada no País, será promulgada pela Mesa da Assembleia, depois de assinada pelos Deputados presentes, e entrará em vigor na data de sua publicação."
42 "Conclusões complementares do Acordo de 1931."

In 1971, based on the joint decision by the Brazilian Academy of Letters and the Academy of Sciences of Lisbon, the National Congress approved Law 5.765, on December 18th of 1971, which introduced three modifications to our orthographic system:

a) the elimination of differential accents on non-homophonous homographs (for example: *gôsto*/*gosto* 'taste/I like'; *bôlo*/*bolo* 'cake/I conceive'; *êle*/*ele* 'he/(the letter) l');
b) the elimination of the indicative accent on the subtonic syllable in words formed with the suffix *-mente* or with suffixes beginning with *z* (for example: *sòmente* 'only'; *cafèzal* 'coffee plantation'; *cafèzinho* 'coffee'; *pèzinho* 'foot'; *amàvelmente* 'kindly');
c) the suppression of the umlaut in unstressed hiatus (for example: *vaïdade* 'vanity'; *saüdade* 'longing').

The differences between the two orthographies are not substantial and do not prevent understanding texts written in one or the other. However, the orthographies of Portuguese, differently from English, for example, are official. This orthographic duplication complicates the international spread of Portuguese, to the extent that documents from international organizations that adopt Portuguese as an official language would need to be duplicated, since they would have to be published in both orthographies (the minister of Culture declared to the Brazil Agency, on November 16th of 2008, that the United Nations had been resisting the adoption of Portuguese as an official language, because it 'did not have a common orthography'[43] and, therefore, the unification agreement 'enables this demand to be met by the international community'[44]); the certification of proficiency in the Portuguese language could not be unified; teaching materials and linguistic resources, such as dictionaries and grammars, produced in one orthography would not work for countries that adopt another; frequently, foreign students of Portuguese have a Brazilian and a Portuguese professor and each one would teach a different orthography; information searches through different technologies would require knowledge of the two orthographies, and so on. To end this odd situation, the Lusophone countries signed an orthographic agreement in 1990, in Lisbon. It was stipulated that it would go into effect on January 1st of 1994, after its ratification by the different nation-states. Since it was not ratified, as predicted, it could not go into effect and, in 2004, it was agreed that it would go into effect after being ratified by three of the eight countries. Despite the opposing reactions in the different countries, especially in Portugal, all of the Portuguese-language countries had ratified the agreement, with the exception of Angola and Mozambique, where the ratification process is ongoing.

43 "não tinha uma ortografia comum."

44 "possibilita essa demanda ser atendida pela comunidade internacional."

The changes are minimal. For Brazilians, the umlaut and some accents were eliminated. For all others, the main modification was the removal of mute consonants, that is, those that do not correspond to any phonological element, which were used for etymological reasons. In the two orthographies, differential accents in a few paroxytone words were removed, which had previously avoided elimination in the beginning of the 1970s.

This agreement is not a reform, but an orthographic unification. If observed from a strictly orthographic point of view, it has gaps and ambiguities. The problem of the hyphen, for example, is still not entirely resolved. However, the importance of the agreement is political: on the one hand, it contributes to establishing the foundations of the Community of Portuguese-Language Countries; on the other, it eliminates the duplication of official orthographies (Faraco 2016, 197).

The *Vocabulário ortográfico comum da Língua Portuguesa*, 'Common orthographic vocabulary of the Portuguese Language'—VOC, provided in the agreement of 1990, is in the process of being prepared, and a part is already available on the platform of the International Portuguese Language Institute (iilp.cplp.org, last accessed 18.02.2022). The Orthographic vocabularies of Angola, Guinea-Bissau and São Tomé e Príncipe still need to be included. The VOC includes not only the vocabulary common to all the national varieties of the Portuguese language, but also the specific vocabulary of each of these varieties (Faraco 2016, 196–197). It will serve as the foundation for the elaboration of a general dictionary of the Portuguese language.

6 Brazil and the Community of Portuguese-Language Countries (CPLP: Comunidade dos Países de Língua Portuguesa)

The political goal of Portugal had always been to create a political entity larger than the territory that it occupied in Europe. During the glorious period of major discoveries, the goal was economic: to control maritime trade. However, this plan was cloaked by an ideological purpose: to spread the Catholic faith. In *Os Lusíadas*, 'The Lusiads', Camões states that it intends to sing 'the glorious memories / Of those kings who were expanding, / Faith and the Empire' (I, 2, 1–3).[45]

During the period of decadence, the dream of future grandeur became even greater. Father Antônio Vieira writes, in the 17th century, after Spanish domination, a prophetic narrative of glories from Portugal, entitled *História do futuro* (2005), 'History of the Future'. In this text, the author aims to rekindle the myth of the Fifth Empire, which comes after the four empires of Ancient Times: Assyria, Persia, Greece and

45 "as memórias gloriosas/ Daqueles reis que foram dilatando,/ A Fé e o Império".

Rome. This empire is Portuguese and it is responsible for taking the Christian faith to all parts of the world. It would be the Empire of Christ.

In the beginning of the 20th century, when Portugal felt humiliated by the English ultimatum of 1890, which demanded that Portugal remove its military forces from the territory between the colonies of Mozambique and Angola, Fernando Pessoa raises the theory of the Fifth Empire. However, this Empire is cultural (Pessoa 2011). Portugal and Brazil would be a single spiritual nation, because both peoples speak Portuguese. This Empire would then be the Empire of the Portuguese Language. Pessoa had previously written, through his heteronym Bernardo Soares: 'My country is the Portuguese language' (1982, 16).[46] Language, in Fernando Pessoa, is not considered a heterogeneous reality, but an entity that exists out of History and space.

In the 1950s, Agostinho da Silva, a Portuguese intellectual, theorized about the Fifth Empire, which would be a spiritual empire embodied in a community of peoples with the Portuguese language and culture. He was the inspiration for the *Comunidade dos Países de Língua Portuguesa* 'Community of Portuguese-Language Countries'—CPLP (Faraco 2016, 242). The Portuguese language is the instrument for spreading ideas of peace, freedom and fraternity in the new Empire.

José Aparecido de Oliveira, Minister of Culture for the José Sarney Administration, organized the First Meeting of the Heads of State and Government of Portuguese-Language Countries in 1989, in São Luis do Maranhão. In this meeting, the *Instituto Internacional de Língua Portuguesa,* 'International Portuguese Language Institute'—IILP, was created to spread and promote the common language of these countries.

After some preparation, the Summit of the Heads of State and Government, which created the Community of Portuguese-Language Countries, an entity that includes Angola, Brazil, Cape Verde, Guinea-Bissau, Mozambique, Portugal and São Tomé e Príncipe, was held on the July 17th of 1996 in Lisbon. Six years later, on May 20th of 2002, East Timor became the eighth member country upon gaining independence. In 2014, Equatorial Guinea became the ninth member of the Community. The general objectives of the CPLP are:

> 'political-diplomatic collaboration among its member-states, namely to reinforce its presence in the international landscape; cooperation in all domains, including education, health, science and technology, defense, agriculture, public administration, communications, justice, public safety, culture, sports and social communication; the creation of projects to promote and spread the Portuguese language' (cplp.org, last accessed 18.02.2022).[47]

46 "Minha pátria é a língua portuguesa."

47 "a concertação político-diplomática entre seus estados membros, nomeadamente para o reforço da sua presença no cenário internacional; a cooperação em todos os domínios, inclusive os da educação, saúde, ciência e tecnologia, defesa, agricultura, administração pública, comunicações, justiça, segurança pública, cultura, desporto e comunicação social; a materialização de projetos de promoção e difusão da língua portuguesa".

With the creation of the CPLP, a supranational organization characterized by a common language, the aim was to coin Lusophony, a term that emerged between the end of the 1980s and the beginning of the 1990s.

With his concept of Lusotropicalism, Gilberto Freyre is one of the theorists behind the idea of Lusophony. For him, everyone who speaks the Portuguese language, 'are one people'.[48] We should be grouped into a federation, which will be the bloc of 'Lusotropical civilization' (2010, 226).[49] The Portuguese carried out a different colonization, because it was romantic, mixed, and tolerant. The Portuguese language is also mixed. Due to this conception, he criticizes defenders of Lusitanian standards and praises diversity. The Portuguese language was valued for its national and subnational varieties. Miscegenation is the foundation of the Lusotropical culture created by the Portuguese, because it is through mixture that this culture is embodied. The fascist regime of Oliveira Salazar appropriated the Lusotropicalism theses of Gilberto Freyre and defended them vehemently. The CPLP is thus a joining of nine countries that share historical, cultural and linguistic elements.

As seen in the history outlined above, the CPLP was always seen as a strategic project by Portugal, but never a priority for Brazilian foreign policy. While the topic of Lusophony is ongoing in Portugal, it rarely appears in Brazil (in the press, for example). The Brazilian priority is South American integration and the development of South-South relations. In spreading Portuguese, Brazil's priority is bilingual education programs on the borders, agreements with South American countries for teacher training, the foundation of UNILA (*Universidade Federal de Integração Latino-Americana* 'Federal University of Latin American Integration'), based in Foz do Iguaçu. For 'sentimental reasons'[50], Brazil does not refuse to participate in the Community, but is not very involved in it, since it does not allow the country to have clear strategic advantages. After all, few political-diplomatic or economic-social objectives would be implemented through the CPLP (Faraco 2016, 304–307). A common language has little practical effect on the formation of political-economic blocs that are meaningful for the world: the European Union, MERCOSUL, NAFTA, etc.

During the Lula administration, there were some initiatives to strengthen Brazil's participation in the CPLP, such as, for example, the creation of a diplomatic mission accredited in CPLP; the advancement of the ratification process and implementation of the orthographic agreement; the I International Conference on the Future of the Portuguese Language held in Brasília; the foundation of UNILAB (*Universidade da Integração Internacional da Lusofonia Afro-Brasileira* 'University of the International Integration of Afro-Brazilian Lusophony'), based in Redenção, Ceará. The Dilma Rousseff administration abandoned this effort, since the President did not participate in

48 "somos um povo só."
49 "civilização lusotropical."
50 "razões sentimentais."

the Summit Meetings and stopped paying the contributions due to the IILP (Faraco 2016, 308).

On the other hand, the issue of Lusophony does not find much interest in Portuguese-Language African countries or in East Timor. After all, Portugal has a low investment capacity to engage these countries in its own strategic project.

The probability of the CPLP becoming a significant international organization is not very encouraging. The political-diplomatic collaboration carries little weight. Cooperation is limited, since the countries are economically very diverse and even those that are better off are not likely to invest (Faraco 2016, 308). On the other hand, the CPLP is not an area of free-trade or economic integration nor will it be, since each member-state is linked to different blocs for this purpose: Portugal is linked to the European Union; Brazil, to MERCOSUL and so on. It is not and will not be a space for people to freely circulate, since the European Union imposes increasingly more restrictions on the acceptance of overseas citizens.

It is up to the CPLP to promote the Portuguese language. However, even the foundation of a common language seems to be insufficient to make the CPLP effective, since Portuguese-language African countries and East Timor present a wide linguistic variety and Portuguese, despite its important political and social functions, is not a language spoken by the entire population there. Brazil and Portugal do not always promote the language in a concerted action.

Therefore, there are also no more concrete actions in this domain, despite having signed important documents on the topic:

Resolução sobre a Promoção e Difusão da Língua Portuguesa, 'Resolution on the Promotion and Spread of the Portuguese Language', adopted in the IX Ordinary Meeting of the Council of Ministers, held in São Tomé in 2004;

Declaração sobre a Língua Portuguesa, 'Declaration on the Portuguese Language', signed in the VII Conference of the Heads of State and Government in Lisbon in 2008;

Declaração da Praia sobre a Projeção da Língua Portuguesa no Mundo, 'Praia Declaration on the Projection of the Portuguese Language in the World', signed in the XIV Ordinary Meeting of the Council of Ministers, held in Praia city in 2009.

The Constitutive Declaration of CPLP states that the Portuguese language is the 'privileged means for spreading cultural creation among peoples who speak it and for the international projection of cultural values, in an open and universalist perspective'.[51] It recognizes that the language is an 'instrument of communication and work in international organizations and enables each of the Countries, in their own regional

51 "meio privilegiado de difusão da criação cultural entre os povos que a falam e de projeção internacional dos valores culturais, numa perspetiva aberta e universalista."

context, to interpret interests and aspirations that are common to everyone' (cplp.org, last accessed 18.02.2022).[52]

In the VI Extraordinary Meeting of the Ministers of the Community of Portuguese-Language Countries, held in Brasília on March 31st of 2010, a resolution was approved, which recommended that the VIII Conference of Heads of State and Governments of the CPLP approve the Plan of Action to Promote, Spread and Project the Portuguese Language. This conference took place in Luanda on July 23rd of 2010 and approved the plan as it was proposed in Brasília. This plan of action was formulated at the International Conference on the Future of the Portuguese Language in the Global System, held in Brasília, from March 25th to 30th of 2010.

The document has six parts: 1) strategies for establishing the Portuguese language in international organizations; 2) strategies for promoting and spreading the teaching of the Portuguese language: (i) teaching the Portuguese language in the space of the CPLP; (ii) strengthening the teaching of Portuguese as a foreign language; 3) the status of developing the orthographic agreement; 4) the public spread of the Portuguese language; 5) the importance of the Portuguese language in the diaspora; 6) the participation of civil society in establishing the plan.

The different recommendations have several levels of generality. They go from very precise suggestions ('To continue the initiative of adopting Portuguese as a working language in the General Conference of UNESCO'[53]) to those with a very broad generality (for example, 'To broaden training options in the Portuguese language in primary and secondary education in non-member countries'[54]).

The Plan of Action for the Promotion and Spread of the Portuguese Language is an instrument for establishing the foundation for working on this subject. They established

> 'the priorities of implementing the Portuguese language in international organizations; the promotion of the Portuguese language, namely through teaching the language in the space of CPLP and its strengthening as a foreign language; the implementation of the Orthographic Agreement, which privileges the existence of scientific and technical terminologies standardized in the entire space of the CPLP; public diffusion through the production and dissemination of audiovisual content in the Portuguese language; the importance and specificity of the diaspora, which are often the ambassadors of the Portuguese language around the world; and the participation of the civil society, in establishing political goals' (cplp.org; last accessed 18.02.2022).[55]

---

52 "instrumento de comunicação e de trabalho nas organizações internacionais e permite a cada um dos Países, no contexto regional próprio, ser o intérprete de interesses e aspirações que a todos são comuns."

53 "Dar prosseguimento à iniciativa de adoção do português como língua de trabalho na Conferência Geral da UNESCO."

54 "Ampliar a oferta de formação em língua portuguesa no ensino básico e secundário em países terceiros."

55 "as prioridades de implementação da língua portuguesa nas organizações internacionais; a promoção da língua portuguesa, nomeadamente através do ensino da língua no espaço da CPLP e do seu

The X Conferences of Heads of State and Governments of the CPLP, held in Dili, on July 23rd of 2014, adopted the Plan of Action of Lisbon, which resulted in the II International Conference on the Future of the Portuguese Language, which took place in 2013, in Lisbon. This plan centered on five axes:

1) the Portuguese language in scientific development and innovation;
2) the Portuguese language in reinforcing entrepreneurship and the creative economy;
3) the Portuguese language in the cooperation among CPLP countries and in communities of the diaspora;
4) the Portuguese language in international organizations;
5) the Portuguese language in teaching speakers of other languages.

Despite these Plans, few concrete actions were taken, which could have a lasting effect on the diffusion of Portuguese: a) to aid in effectively establishing quality teaching of Portuguese in Latin American countries; b) to aid in the universalization of learning Portuguese in African and Asian countries, in which it is the official language; c) to increase the number of lecturers through a system of counterparts in foreign universities; d) to implement a system for translating significant works, mainly literary ones, from Portuguese-language cultures to languages with a significant number of speakers; e) to significantly increase the number of grants for training university professors of Portuguese; f) to create CPLP TV according to the model of TV5 Monde. Until today, not much progress has been made in establishing the Digital Library of CPLP and the Portal of the Portuguese Language is still poor.

The International Portuguese Language Institute is the main instrument for establishing the policy of promoting and spreading Portuguese. However, it has experienced serious financial problems, because the countries in the community do not regularly pay the contributions owed, which reveals a lack of high engagement among the member-states in administrating actions regarding the language. Despite this, it has taken on some important projects, such as the elaboration of the Common Orthographic Vocabulary of the Portuguese Language, and is taking the first steps towards elaborating Technical and Scientific Terminologies in the Portuguese Language.

The truth is that there are no policies converging to increase the space of Portuguese among international languages. Portugal and Brazil often develop specific and independent policies. For example, there is still no system for mutually recognizing certificates of proficiency.

---

fortalecimento como língua estrangeira; a implementação do Acordo Ortográfico, que privilegia a existência de terminologias científicas e técnicas harmonizadas em todo o espaço da CPLP; a difusão pública, através da produção e disseminação de conteúdos audiovisuais em língua portuguesa; a importância e especificidade das diásporas, que são muitas vezes os embaixadores da língua portuguesa pelo mundo fora; e a participação da sociedade civil, na concretização das metas políticas."

Brazil has around 85 % of the Portuguese speakers and, therefore, should have a central role in the actions that address the future of the language and its diffusion. However, there is no broad strategic vision on this issue. Transforming the language into an economic asset has not been considered. Therefore, actions for promoting the language are very limited. Moreover, these actions are taken by three ministries: Foreign Affairs, Education, and Culture. This generates a certain political dispersion and indeterminacy. Contrary to Portugal, whose actions for promoting and spreading the language are centered on the Camões Institute, Brazil lacks an organization that coordinates this activity. Brazil would need a broader and more cohesive plan in order to promote and spread Portuguese.

7 References

Alencar, José de (1965), *Obra completa*, 3 vol., Rio de Janeiro, Aguilar.

Alencar, José de ([19]1995), *O guarani*, São Paulo, Ática.

Almeida, Rita Heloísa de (1997), *O Diretório dos índios: um projeto de 'civilização' no Brasil do século XVIII*, Brasília, Editora de UNB.

Anchieta, José de ([7]1990), *Arte de gramática da língua mais usada na costa do Brasil*, Coimbra/São Paulo, Loyola.

Andrade, Mário de (1965), *Macunaíma: o herói sem nenhum caráter*, São Paulo, Martins.

Camões, Luís (1988), *Obra completa*, Rio de Janeiro, Nova Aguilar.

Dias, Luiz Francisco (2001), *O nome da língua no Brasil: uma questão polêmica*, in: Eni Pulcinelli Orlandi (ed.), *História das ideias linguísticas: construção do saber metalinguístico e constituição da língua nacional*, Campinas/Cáceres, Pontes/Unemat Editora, 185–198.

Faraco, Carlos Alberto (2016), *História sociopolítica da língua portuguesa*, São Paulo, Parábola.

Freyre, Gilberto (2010), *Um brasileiro em terras portuguesas*, São Paulo, É Realizações.

Herder, Johann Gottfried (1987, [1]1772), *Ensaio sobre a origem da linguagem*, Lisboa, Edições Antígona.

Herder, Johann Gottfried (1996), *Sur la nouvelle littérature allemande. Fragment*, *Lettres sur l'avancement de l'humanité*, in: Pierre Caussat/Dariusz Adamsky/Marc Crépon, *La langue source de la nation. Messianismes séculiers en Europe centrale et orientale (du XVIIIe au XXe siècle)*, Liège/Bruxelles, Mardaga, 77–106.

Holanda, Sérgio Buarque ([3]1956), *Raízes do Brasil*, Rio de Janeiro, José Olympio.

Lucchesi, Dante (2015), *Língua e sociedade partidas: a polarização sociolinguística no Brasil*, São Paulo, Contexto.

MEC (Ministério da Educação e Cultura) (1997), *Parâmetros curriculares nacionais: língua portuguesa*, Brasília, MEC.

Melo, Gladstone Chaves de ([3]1972), *Alencar e a língua brasileira*, Rio de Janeiro, Conselho Federal de Cultura.

Payer, Maria Onice (2001), *A interdição da língua dos imigrantes (italianos) no Brasil: condições, modos, consequências*, in: Eni Pulcinelli Orlandi (ed.), *História das ideias linguísticas: construção do saber metalinguístico e constituição da língua nacional*, Campinas/Cáceres, Pontes/Unemat Editora, 235–256.

Pereira, Eduardo Carlos ([110]1958), *Gramática expositiva: curso superior*, São Paulo, Companhia Editora Nacional.

Pessoa, Fernando (2011), *Sebastianismo e Quinto Império*, Lisboa, Ática.
Pinto, Edith Pimentel (1978), *O português do Brasil: textos críticos e teóricos*, vol. 1: *1820–1920*, Rio de Janeiro/São Paulo, Livros Técnicos e Científicos/Editora da Universidade de São Paulo.
Ramos, Heloísa (2009), *Por uma vida melhor*, São Paulo, Global.
Rodrigues, Aryon (1993), *Línguas indígenas: 500 anos de descobertas e perdas*, DELTA—Revista de documentação em linguística teórica e aplicada 9/1, 83–103.
Saussure, Ferdinand de (1969), *Curso de linguística geral*, São Paulo, Cultrix/EDUSP.
Silva Neto, Serafim da (1977), *Introdução ao estudo da língua portuguesa no Brasil*, Rio de Janeiro, Presença.
Silveira, Souza da (1960), *Lições de português*, Rio de Janeiro, Livros de Portugal.
Soares, Bernardo (1982), *Livro do desassossego*, vol. 1, ed. Jacinto do Prado Coelho, Lisboa, Ática.
Viana, Aniceto Reis Gonçalves/Abreu, Guilherme de Vasconcellos (1885), *Bases da ortografia portuguesa*, Lisboa, Imprensa Nacional.
Viana, Aniceto Reis Gonçalves (1904), *Ortografia Nacional. Simplificação e uniformização sistemática das ortografias portuguesas*, Lisboa, Editora Viúva Tavares Cardoso.
Vieira, Antônio (1959), *Sermões*, vol. 1, t. II, Porto, Lello & Irmão.
Vieira, Antônio (2005), *História do futuro*, Brasília, Editora da UnB.
Williams, Edwin B. (1975), *Do latim ao português*, Rio de Janeiro, Tempo Brasileiro.

Elisângela N. Teixeira and Michele C. dos S. Alves

20 Psycholinguistic studies: language acquisition and processing

Abstract: This chapter presents an overview of psycholinguistic studies that have investigated the processing and acquisition of Brazilian Portuguese linguistic structures. It is divided into five sections. In the first section, we present the criteria used to select the studies that we will discuss in the following sections. In addition, we briefly introduce the reader to the field of psycholinguistics. In the second section, we present the experimental methods most commonly used in the field today. The third section is organized into two topics: the first of these discusses some of the contributions in the field of language acquisition that have investigated aspects such as the acquisition of gender, number, verbal affixes, and relative clauses; in the second, the studies presented in the field of language processing deal with coreference processing, agreement, and relative clauses. All discussions are based on experimental research papers with stimuli materials in Brazilian Portuguese. To conclude, the fourth section describes the history of the development of the discipline in Brazil and some of the most notable Brazilian centers of psycholinguistic research, and it is followed by some final remarks.

Keywords: psycholinguistics, language acquisition, language processing, syntactic processing, Brazilian Portuguese

1 Introduction

It is commonplace to define psycholinguistics as the junction between linguistics and psychology, and beyond that, as a discipline that binds language acquisition and language processing studies. However, it is not simple to delineate a boundary between psycholinguistics and linguistics, in the sense that almost every linguist in the world wants to understand the mechanisms that underlie our ability to communicate by sign, oral, and written language.

"Psycholinguistics is the study of the mental processes and skills underlying the production and comprehension of the language, and of the acquisition of these skills" (Levelt 1992). Psycholinguistics aims to describe the underlying processes related to our capacity to communicate through a shared language in a certain community, rather than to generate knowledge about any particular language, such as Brazilian Portuguese (BP), for example.

Psycholinguistics first emerged after two summer seminars at Cornell University in 1951 and at Indiana University in 1953 (Levelt 2013). The studies presented at those

https://doi.org/10.1515/9783110405958-021

seminars were subsequently published in *Psycholinguistics: a survey of theory and research problems* by Osgood/Sebeok (1954), thus marking the beginning of the field. By the end of 1950s, the publication of *Syntactic Structures* (Chomsky 1957) added to the new discipline of psycholinguistics, creating a fertile atmosphere for innovative research ideas and methodologies—a period that has become known in the history of linguistics as the "cognitive revolution".

It is important to note that at the beginning of 2021, more than ever, psycholinguists are seeking to resolve, through the use of computational methods and big data (Manning et al. 2020), the controversial debate on the role of the Faculty of Language in the narrow-sense (Hauser/Chomsky/Fitch 2002) and the role of linguistic input in language acquisition. In other words, psycholinguists are trying to identify what kinds of mechanisms allow humans to infer regularities from a large corpus of propositions received during the development of their languages, and are doing so through conducting experiments using behavioral, neurophysiological and computational methods. Psycholinguistics is considered to be one of the areas of cognitive science most closely related to cognitive psychology, the neuroscience of language, artificial intelligence, and computational linguistics. After many years of work, it is a dynamic field which is continuously creating new areas of study, such as biolinguistics, which studies language within the framework of evolution by describing how human biology constrains the lexicon and language structure, and socially-situated language processing studies, which investigate how social aspects influence the comprehension and production of speakers' utterances.

This chapter aims to provide an overview of some of the most relevant contributions of experimental research in psycholinguistics conducted in Brazil. The chapter is structured into five sections: "Introduction"; "Most common on-line and off-line methodologies"; "Psycholinguistics studies", this divided into two subsections, "Language acquisition" and "Language processing studies"; "The establishment of psycholinguistics at Brazilian universities", which provides a brief summary of how the field was developed in Brazil, followed by "Final remarks".

To be as fair as possible in representing the legacy of the field, we have established some criteria. First, we listed active psycholinguistics researchers in Brazil, most of them members of the psycholinguistics group *Associação nacional de pós-graduação e pesquisa em letras e linguística,* 'National association for research and graduate studies in letters and linguistics'—ANPOLL. Then, we selected the most-cited experimental research articles of these authors using the *Google Scholar* database. From the resulting list of 134 articles, we selected a subset of articles whose objects of study were related to some linguistic structures of BP. Review papers and popular science articles were excluded from the selection, as were papers that compared languages, studied language disorders, investigated bilingualism or multilingualism, or described aspects of acquisition, development, or processing of reading, since the focus of this manual is on BP. Research papers on the acquisition of BP phonology were also excluded since the reader can find more information on this topic in another of the

chapters in this manual (↗12 Sound-related aspects of Brazilian Portuguese). Thus, we selected a total of 53 articles, 11 dedicated to language acquisition and 42 to language processing.

We then grouped these articles by theme, dividing them into two subareas: language acquisition and language processing. Among the language acquisition articles, we found six subjects: agreement, relative clauses, prepositional phrases, verbal and nominal affixes, functional categories, and numerical cognition. Then we summed the total number of citations by subject, this to identify the three most frequently studied and cited subjects, which were: the acquisition of gender and number, verbal affixes, and relative clauses. Among the language processing research papers, we found eleven different subjects: coreference processing, agreement, relative clauses, lexical access, syntax and prosody interface, temporary and permanent ambiguity, verbal aspect, countable nouns processing, syntactic islands, topic, and determiner phrases. Using the same procedure, from the sum of the total number of citations per subject, we selected the three most-cited topics: coreference, agreement, and relative clauses.

2 Most common online and offline methodologies

Currently, psycholinguists use a diverse range of experimental methods, which can be classified into behavioral and neurophysiological approaches. Behavioral methods, in turn, can be divided into offline and online methods. In this section, we will present the most common methods used in psycholinguistics, some of which were used in the Brazilian research papers discussed in section 3.

Behavioral methods measure participants' behavioral reactions to the independent variables controlled for in experiments. Metrics are often the response latencies, that is, the time it takes a participant to react to an experimental item of a particular experimental task. One of the important assumptions of this type of research is that behavioral measures reflect cognitive difficulty.

Offline methods differ from online ones in that the latter measure behavioral or neurophysiological responses during linguistic processing, while the former measure behavioral responses after processing, that is, responses that involve a greater degree of awareness by the participant.

Eye tracking is an online behavioral method that records automatic and unconscious eye movements during a task, such as reading a written text or listening to speech, this usually matched with a picture (see visual world paradigm in Tanenhaus et al. 1995). As examples of neurophysiological methods, the three techniques most commonly used by psycholinguistics are those which measure the event-related potentials (ERPs) by capturing electrical activity of the scalp through electroencephalography (EEG), changes in magnetic field recorded by magnetoencephalography (MEG), and changes in brain blood flow recorded by the functional magnetic resonance image (FMRI).

Offline methodologies are widely used. Among various kinds of questionnaire-based methods, the most common are tasks using acceptability or preference judgments (formerly called grammaticality judgments). Acceptability judgment consists of presenting a list of grammatical and ungrammatical sentences to participants, whose task is to rate the sentences, using yes-no answers, forced choices, a Likert scale, or with magnitude estimation, among others. The assumption of such offline testing is that judgments are a way of accessing the speaker's internal grammar. The term grammaticality has now been replaced by acceptability or preference, since grammaticality is a difficult concept for participants unfamiliar with the science of language. It is important to note that this is a widely used method, and not exclusive to psycholinguistics.

Other offline methods include cloze tests, priming or cross-modal priming tests, truth value judgment tasks, sentence-picture match tasks, act-out procedure tasks, the wug test (Berko 1958) and self-paced reading tasks (a technique that could be considered to have a high online resolution if the sentence/text is presented in a non-cumulative word-by-word procedure). In general, these offline tests seek to assess whether the sentences or words (targets) presented are well understood by a group of participants.

One of the techniques mostly used to investigate language acquisition is the acting-out procedure, which consists of presenting a story to babies or young children and asking them to act-out with dolls what they have just understood. This technique is very simple and feels like a game to participants, which facilitates the study of the limitations on the interpretation of certain linguistic structures in acquisition (see 3.1.3).

Another popular online research methodology in the language acquisition of babies is preferential looking (Fantz 1961). Seated in front of a screen, an image book, or a puppet show, babies are presented with visual/auditory stimuli for a certain period of time, which is called the familiarization phase. Then, in the testing phase, they are presented with congruent and incongruent trials while a camera nearby records their looks and in some cases their pupil dilation (if eye trackers are used). The assumption of this method is that children look in different ways when they hear ungrammatical sentences or some other incongruent stimuli. Furthermore, children look more attentively at stimuli which are well-formed or in some other way are preferential to them. In other words, the preferential looking methodology captures the babies' attention through tracking the direction of their eye movements (see section 3.1).

For experiments with older children, among several methods, the oral production method is one of the most common. Here children are exposed to visual/auditory stimuli on a screen or image book. During the familiarization phase, they can learn, for example, novel words, or even watch/listen to a story. In the testing phase, they are required to give a certain response, this generally elicited by a character on the video or a puppet asking a question. Their responses are recorded by the experimenters (see section 3.1). Instead of answering a question, children can also be instructed to point to an image or toy (cf. Corrêa/Name 2003).

For language processing, priming experiments are also widely used. These can be associated, for example, with a lexical decision task in studies on lexical processing (cf. Corrêa/Almeida/Porto 2004) or a probe recognition task (cf. Leitão 2005) in studies on sentence processing. The priming effect in psychology is known as a facilitatory subconscious response caused by a previous stimulus. Psycholinguists use priming experiments to investigate the organization of the mental lexicon, or the strength of activation of certain lexical units or structural constructions. Priming stimuli can be visually presented with either text or image materials on a computer screen, or they can mingle visual and audio stimuli, this is known as cross-modal priming task (Swinney 1979). In visual priming experiments with a lexical decision task, participants are presented with an initial cross centered on the screen, followed by primes, which are words subliminally presented on the screen for a very short period of time. Primes can then be followed by either another centered cross on the screen or by the target stimuli. In this case, participants are instructed to decide whether the targets are words by pressing a button on the computer keyboard.

Experimenters can manipulate the time the participants are exposed to the primes and the targets to capture early or late processing effects or to examine a myriad of language phenomena involving phonological, semantic, or syntactic associations between words. In priming experiments with a probe recognition task, participants are presented with sentences on a computer screen, and are then instructed to decide whether a word appears or not in the sentence. Experimenters can manipulate when targets appear, be it at the end of each sentence or at a specific syntactic position within it.

Another sentence processing method is self-paced reading task (cf. Leitão 2005; Almor et al. 2017; Oliveira/Leitão/Araújo 2013; Leitão/Peixoto/Santos 2008). Here the reader's pace is recorded while words or phrases of a sentence/short text are revealed by pressing the keyboard spacebar. The experiment stimuli can be presented in a cumulative or non-cumulative way, that is, revealed words or portions of the sentence/short text remain on the screen (cumulative) or are hidden again (non-cumulative). The task of the participant is to read each segment of a sentence or short text at their own pace. The response time it takes for a participant to press the spacebar (for example) is recorded in milliseconds and corresponds to the reading time for each segment. The assumption is that the more complex language structures are, the longer they will take to be processed by the human sentence processor (parser), resulting in longer reading times. Generally, at the end of each sentence, participants are given comprehension questions in order to check their attention during the task.

If the self-paced reading task is presented word-by-word in a non-cumulative way, the response time is very reliable, and for this reason it is considered an online method. Many results found in experiments using self-paced reading tasks have been confirmed by finer-grained online methods such as eye tracking. Eye tracking is more fine-grained than self-paced reading because it can involve different measures over the course of language processing, from early to late processing stages, including rea-

nalysis. Eye trackers can record the participants' eye movements, and the two most studied eye movements that can be captured relating to reading studies and language processing are fixations and saccades. Fixations are the duration of time that the eyes remain on a certain part of the text/image, and saccades are the movements our eyes make moving forwards (around 90 % of the time in skilled readers) and backwards (around 10 % of the time in skilled readers) through a text or image, for example (Staub/Rayner 2007).

Several linguistic processes can be studied using the methods described above. In the next section, we will describe how some of these methods, such as oral production tasks, truth value judgment tasks, the acting-out procedure, self-paced reading, and eye tracking, among others, have been used to investigate agreement, the acquisition of verbal affixes, anaphora resolution, and the acquisition and processing of relative clauses.

3 Psycholinguistics studies

3.1 Language acquisition studies

3.1.1 Morphological acquisition: gender and number

As a Romance language, BP attracts linguistic studies on agreement since it has redundant and visible morphological marking. Corrêa/Name (2003) investigated how children acquire the gender system in BP. In Portuguese, all nouns have gender, masculine or feminine. The gender of animate nouns is associated with the sex of their referents, with very few exceptions, whereas inanimate nouns have arbitrary gender. Feminine gender is morphologically marked by an *-a* ending while masculine gender, which is the default gender, is morphologically marked by an *-o* ending.[1] Besides that, gender is redundantly marked in determiners, adjectives, and participial forms.

Corrêa/Name (2003) was interested in examining how children discover the gender of the words in Portuguese in order to comprehend and produce grammatical utterances. The authors' hypothesis is that gender is a shared feature within the Determiner Phrase (DP), allowing the gender of the determiner to be assigned to the noun in order to provide agreement within the DP. Corrêa/Name (2003) conducted two experiments, testing children's comprehension and production, respectively.

In the comprehension experiment, 32 children who had been acquiring BP for between 21 and 28 months were presented with auditory stimuli. Specifically, a puppet called Dedé asked the children to point to a picture of an already familiar word.

1 It should be noted that only feminine gender is definitely lexically specified in BP, whereas masculine gender might be underspecified, since it is a default gender (Mattoso Câmara Jr. 1970; Harris 1991).

The independent variable was the determiner used, which could be either congruent with the gender of the noun, as in (1a) below, incongruent to the gender of the noun, as in (1b), or could be substituted by either a Portuguese complementizer (1c) or a pseudo-determiner (1d). Furthermore, there was a control condition with words randomly displayed within the sentence (1e). The dependent variable was the children's responses.

(1) a. Mostre a bola pro Dedé.
'Show the(FEM) ball(FEM) to Dedé.'

b. Mostre o bola pro Dedé.
'Show the(MAS) ball(FEM) to Dedé.'

c. Mostre se bola pro Dedé.
'Show if ball(FEM) to Dedé.'

d. Mostre biu bola pro Dedé.
'Show biu ball(FEM) to Dedé.'

e. Dedé a mostre bola pro.
'Dedé the(FEM) show ball(FEM) to.'

Based on the responses of the 14 children who presented valid responses, the results of the comprehension experiment showed that children tend to answer more correctly in the congruent response (92.21 %), as in (1a), than in the incongruent response (76.64 %), as in (1b). There were no differences between the complementizer (63.64 %) (1c) or the pseudo-determiner (63.64 %) (1d). Thus, children at the age of 2 can perceive morphological marking in determiners in order to assign gender to the nouns and correctly comprehend utterances.

In the production experiment, Corrêa/Name (2003) investigated whether children would be sensitive to feminine and masculine phonological marking in Portuguese. The authors tested two groups of 15 children, one with children younger than 3 years of age (mean age 16 months) and the other with children older than 3 (mean age 54.2 months). The children were shown a sequence of 3 pictures presenting a novel word followed by a determiner with phonological gender marking. The independent variable was the morphological marking of the pseudo-noun, which could correlate either positively (dab*o*(MAS), bid*a*(FEM)) or negatively (dab*o*(FEM), bid*a*(MAS)) with the phonological gender marking of masculine and feminine in Portuguese; or it could be opaque (mip*e*(MAS), mip*e*(FEM)). The dependent variable was that of the children's responses. For example, in the first picture, children would listen to a sentence like (2a); in the second picture, they would listen to a sentence like (2b); and in the third picture, they would listen to a picture like (2c). Then, in the fourth picture they would

be asked to correctly produce the gender of the novel word, as in (2d), and expected responses would be the ones like (2e).

(2) a. Isso aqui é uma depa.
'This here is a(FEM) depa(FEM).'

b. Olha, aqui tem uma depa também.
'Look, here there is a(FEM) depa(FEM) too.'

c. As depas estão juntas no armário.
'The(FEM) depas(FEM) are together in the cupboard.'

d. Oh! Uma depa sumiu! Que depa sumiu?
'Oh! One(FEM) depa(FEM) has disappeared! Which depa(FEM) has disappeared?'

e. A depa vermelha./Essa aqui.
'The(FEM) red(FEM) depa(FEM)./This(FEM) one here.'

The results of the production experiment showed that out of the maximum score of 3, in the positive correlation, children younger than 3 years old scored an average of 2.93, while the older children scored 3.0. In the negative correlation, children younger than 3 years old scored 2.67, while the older children achieved 1.87. Finally, in the opaque condition, children younger than 3 years old scored 2.8, and the older children 2.73. The authors concluded that children younger than 3 years old are sensitive to masculine and feminine phonological patterns in Portuguese, and this sensitivity increases with age. It is important to note, however, that this phonological strategy to discover the gender of the words relies on the morphological cues shown by the determiner.

Corrêa/Name (2003) concluded that the results of both experiments confirmed their hypothesis. The acquisition of the gender system in BP provided evidence that gender agreement works under a feature-sharing grammatical operation within the DP, in which the gender of the determiner is assigned to the noun. Furthermore, the authors provided evidence that children at the age of 3 are not only able to recognize the gender of the determiners, but are also sensitive to the masculine and feminine phonological marking strategy in order to discover the gender of the nouns and in turn to correctly understand utterances and produce grammatical utterances.

Besides the acquisition of gender, Brazilian psycholinguists have also investigated the acquisition of number. In BP, children are exposed to two number systems. In the standard variety, plural is redundantly marked with the inflection *-s* in the noun endings, as well as in all elements within the DP such as determiners, adjectives, and participial verb forms. In the standard variety, plurality is also expressed through sub-

ject-verb agreement and subject-predicate agreement. On the other hand, in the non-standard variety, plurality can only be marked in the determiner or the most leftward element in the DP, and sometimes in the verb (↗10 Sociolinguistics).

Taking this into account, Corrêa/Augusto/Ferrari-Neto (2006) investigated whether children were able to identify plurality in both the standard and non-standard varieties of BP. The authors explained that when presented with the non-standard variety, children would only rely on the plural morphological marking of the determiner to capture the plurality expressed within the phrases or sentences. Thus, their hypothesis was that the number marking in the determiner would enable number assignment to all elements within the DP through a c-command. This strategy could also be very helpful in the acquisition of novel nouns since children would be able to identify a semantic property of a novel noun using the information on plurality expressed in the determiner.

Corrêa/Augusto/Ferrari-Neto (2006) conducted a psycholinguistic experiment with 18 children from 18 to 30 months (mean age 25 months) who were acquiring BP. Children were presented with a puppet called Dedé, who would ask them to point to a picture with multiple tokens of a novel object/creature on the page of a book that contained 4 images. The independent variables were: (i) the plural expression in the DP, which could be grammatical (with the morpheme *-s* expressed in the determiner) or ungrammatical (without the morpheme *-s* expressed in the determiner) and (ii) the variety, either standard (with the morpheme *-s* redundantly marked in the determiner and in the noun) or non-standard (with the morpheme *-s* only marked in the determiner). Moreover, there was a control condition in the singular. The dependent variable was the correct responses given by the children. For example, a grammatical sentence in the plural in the standard variety is illustrated in (3a) below, while the same sentence in the non-standard variety is illustrated in (3b); an ungrammatical sentence is illustrated in (3c).

(3) a. Mostra os dabos pro Dedé.
'Show the(PL) dabos(PL) to Dedé.'

b. Mostra os dabo pro Dedé.
'Show the(PL) dabo(SG) to Dedé.'

c. Mostra o dabos pro Dedé.
'Show the(SG) dabos(PL) to Dedé.'

The results showed that children were able to give more multiple-token responses in the grammatical condition (61.1 %), as in (1a) and (1b), than in the ungrammatical condition (31.4 %), as in (3c). This means that by the age of 18–30 months children already know the number system in BP, that is, they know that plural markings must be expressed in the determiner. However, a statistically significant difference between the

standard (1a) and the non-standard (1b) varieties was not found, which means that children identify information on plurality exclusively though the plural morphological marking expressed in the determiner. In other words, the redundant plural marking in the nouns did not increase the chances of children giving multiple-token responses.

Corrêa/Augusto/Ferrari-Neto (2006) concluded that children at the age of 2 are able to recognize plural morphological number in BP by detecting the morphological marking expressed in the determiner and by establishing number agreement relations between the determiners and the other elements in the DP. This study provided evidence that children were able to use phonological and morphophonological distributional information to work out functional categories such as determiners in order to detect semantic information like plurality. In the literature, this is known as bootstrapping.

Castro/Ferrari-Neto (2007) replicated the experiment reported in Corrêa/Augusto/Ferrari-Neto (2006) with children acquiring European Portuguese (EP). Unlike BP, EP does not have the possibility of marking plural in the DP by exclusively using the morpheme *-s* in the determiner, which is a typical characteristic of the non-standard BP variety. Results indicated that while children acquiring BP did not show any statistically significant differences between (3a) and (3b), children acquiring EP gave more multiple-token responses in (3a) than in (3b), 79.67 % and 46.67 % respectively. Thus, it seems that children acquiring Portuguese (BP or EP) perceive plural marking in the determiner as a plural cue, but they do so at different levels depending on the variety. Furthermore, children acquiring BP made more mistakes, giving more multiple-token responses in the singular condition (12.5 %) than children acquiring EP (4.44 %). These results indicate that the performance of children acquiring EP is closer to the standard variety once its number system is more stable, with only one number marking possibility; this might indicate that, because children acquiring BP are exposed to two competing number systems, the acquisition of morphological number marking may be stabilized later in BP than in EP.

3.1.2 Morphological acquisition: verbal affixes

In this section, we will discuss some research findings on the acquisition of verbal suffixes in BP (Bagetti/Corrêa 2011; Pires/Rothman 2009a). The studies to be presented here explore evidence of language acquisition through experimental research that combines psycholinguistics hypotheses with theoretical concepts of the Minimalist Program (Chomsky 1995).

Previous studies on language acquisition (Galves 2001; Kato 2005; Roberts/Kato 1993) revealed that children would not be able to acquire inflected infinitives in a natural, non-school context because this structure is absent in colloquial varieties of BP. On the same lines, Pires/Rothman (2009b) compared the acquisition of inflected infinites by BP and EP heritage speakers in a judgment/correction task, finding that heri-

tage speakers of BP, unlike the heritage speakers of EP, could not differentiate inflected infinitives from non-inflected ones. Such a difference was explained due to the fact that unlike BP, inflected infinitives are colloquially used in EP.

With the aim of clarifying the situation regarding the acquisition of inflected infinitives in BP, Pires/Rothman (2009a) designed a cross-sectional study with children from 6 to 15 years in order to identify more precisely when children acquiring BP acquired or parameterized the syntactic and semantic characteristics of inflected infinites. This was the first study on the acquisition of inflected infinitives by monolinguals and native speakers of BP. The authors were also interested in understanding the role of free acquisition and the role of learning in a formal school context on this process, and did so by using test sentences such as (4).

(4) a. Eu_i convenci o Leo_j $PRO_{j/*i+j}$ a dividir o sorvete com o João.
I_i convinced the Leo_j $PRO_{j/*i+j}$ to share-INF the ice-cream with the João
'I convinced Leo to share the ice cream with João.'

b. Eu_i convenci o Leo_j $pro_{i+j/i+j+k}$ a dividirmos o sorvete com o João.
I_i convinced the Leo_j $pro_{i+j/i+j+k}$ to share-INF-1PL the ice-cream with the João
'I convinced Leo for us to share our ice cream with João.'

Pires/Rothman (2009a) distributed 87 children and teenagers, all native speakers of BP, into 4 age groups: 6–7, 8–9, 10–12 and 13–15 years. The authors conducted two experiments: a Morphological Recognition Task (MRT) and a Context Match Task (CMT). The materials involved stories with toys for children up to 9, and written stimuli for children 10 and over. The children and teenagers were told that the characters were still learning Portuguese and that they would need help to speak the language; therefore, the participants were asked to correct the sentences uttered by the characters. The stories contained non-inflected infinitives and inflected infinitives in the present.

In the MRT, it was found that participants in all the age ranges could differentiate the correct and incorrect use of inflected infinitives, with an increased capacity proportional to age. However, only the children in the 13–15 range were able to recognize the incorrect use of inflected infinitives at an above-average rate, like the adults did. Until 12 years, children do not have complete grammatical competence regarding inflected infinites. However, it should be noted that this task yields passive data, and the ability to recognize is different from production. In the CMT, participants were asked to differentiate the syntactic and the semantic contexts that allowed the use of inflected infinitives in a truth value judgment task. The results showed that under 10 years, children do not have an internalized grammar of the inflected infinitives, since they did not perform at above-average, like adults do here.

Pires/Rothman's (2009a) findings indicated that the grammatical competence of inflected infinitives is a result of exposure to the standard dialect at school rather than

exposure to the colloquial variety. The authors argued that competence in L1 BP can occur in two different ways: as a result of exposure to the colloquial spoken language in its social context, or as a result of formal education, which is the case for older speakers in their youth and early adulthood. A useful method to test late acquisition of inflected infinitives would be to examine online processing in order to analyze the difference in reaction times between superficial and deep processing.

Bagetti (2009) and Bagetti/Corrêa (2011; 2013) conducted experiments investigating the morphophonological patterns of verbal affixes in babies acquiring BP. The authors sought to identify when very young children pass from the sound perception stage (phonological) to the morphosyntactic representation of verbal affixes that encode grammatical information such as tense, aspect, mood, person, and number, as well as subject-verb agreement. The hypothesis was that there would be an early sensitivity to the patterns of verbal suffixes at the end of the first year, which explains why 18-month-old children already process subject-verb agreement, and are also able to perform suffix stripping. This early-acquired skill is also a prerequisite for specifying the other functional categories in BP. In previous studies (see section 3.1.1, above), Name/Corrêa (2003) found that 15-month-old children recognize phonological changes in determiners in fluent speech, showing a significant behavioral reaction to a pattern break of the determiners in BP, which suggests that the acquisition of this functional category begins at that age. Thus in BP, as in other languages (Höhle/Weissenborn 2000; Jusczyk/Luce/Charles-Luce 1994; Shi/Werker/Morgan 1999; Shady 1996; Shafer et al. 1998), very young children can distinguish functional from lexical categories, which indicates that they have a minimum lexicon at the end of their first year of life.

Bagetti/Corrêa (2011) reported two experiments on preferential hearing. The authors recruited 38 babies of between 9 and 16 months and divided them in two groups, 9–12 and 12–16 months. In the experiments reported in Bagetti/Corrêa (2011), auditory and visual stimuli were simultaneously presented, and the authors manipulated nominal affixes (R-MOD) and verbal affixes (AFMOD) by changing some diphthong sounds, as shown in Table 1.

Table 1: Example of an experimental item used in Bagetti/Corrêa (2011).

Story type	Story
Normal	O rei descobriu uma lagoa e estabeleceu uma lei: – É só minha! O plebeu partiu e o besouro não mais se banhou. O hebreu enfureceu-se e mandou secar a lagoa. O rei se arrependeu. O hebreu mandou encher a lagoa.
AFMOD	O rei descobr**ɛy** uma lagoa e estabeles**ɔʀ** uma lei: – É só minha! O plebeu part**ɛy** e o besouro não mais se banh**un**. O hebreu enfures**ɔʀ**-se e mandou secar a lagoa. O rei se arrepend**ɔʀ**. O hebreu mand**un** encher a lagoa.

Table 1: (continued)

Story type	Story
R-MOD	O rɔw descobriu uma lagoa e estabeleceu uma lɔw: – É só minha! O plebɔʀ partiu e o besunʀo não mais se banhou. O hebrɔʀ enfureceu-se e mandou secar a lagoa. O rɔʀ se arrependeu. O hebrɔʀ mandou encher a lagoa.

The difference between the two experiments in Bagetti/Corrêa (2011) was the position in which the affixes were manipulated. In one experiment, the authors manipulated the position independently of the phonological environment, thus affecting the syllabic pattern of BP, while in the other experiment, the affixes were manipulated according to their morphological environment, without changing the syllabic pattern of BP. The dependent variable was the story hearing time, that is, the time the children maintained their head turned (left or right) towards the audio source. When the children's attention was lost for more than 2 seconds as a result of the manipulated factors, such as changes in the verbal or noun suffixes, this was considered to be a loss of attention, indicated by a child's head turning away from the audio source. The authors predicted that the shorter the hearing time (that is, the child's attention), the stronger would be their sensitiveness to the morphophonological representations of the functional verbal categories. In other words, they predicted that the hearing time would not be different between the controlling condition (NORMAL) and the condition with noun changes (R-MOD), but that it would be different in relation to the condition with verbal changes (AFMOD).

Bagetti/Corrêa's (2011) findings revealed that children detected phonic changes that do not change the syllabic pattern of the language in both functional and lexical words, but they detected morphophonological changes, especially in functional elements. In this way, the authors were able to identify when children pass from the phonic level to the morphophonological representation of functional elements. Children between 9 and 12 months old were only sensitive to changes of verbal affixes (AFMOD), while children from 13 to 15 months old seemed to have a more complex lexicon, which was responsible for making them react negatively to changes of the noun suffixes (R-MOD), which indicated a more complete perception of the lexical categories. It seems, then, that the class of functional words might in the first place be determined by the phonic interface of these words, so that they can be specified according to their formal characteristics. Bagetti/Corrêa (2011) can be considered as a pioneer study, since very little work has been dedicated to exploring the acquisition of functional elements in BP. The best-known are the ones from PUC-Rio School (see section 4 in this chapter).

3.1.3 Syntactic acquisition of relative clauses

Comprehension of relative clauses is understood as fundamental to the description of language development since it is a resource that reflects the human language capacity of expanding the adjective function of noun phrases by using embedded sentences.

The traditional method of investigating language acquisition in children's production is through the acting-out procedure, which consists of an experimental task in which children are asked to act out the linguistic stimuli by using toys to represent the subjects (actors), the objects, and the verbal action. For example, in a sentence such as (5), the children would use the toy animals to act out the actions expressed by the main clause and by the relative clause – the horse would kick the cow and the cow would push the sheep.

(5) O cavalo chutou a vaca que empurrou a ovelha.
(actor – action – object – relative marker (actor) – action – object)
'The horse kicked the cow that pushed the sheep.'

Then, through the standard acting-out methodology, which can be found in several studies (Bowerman 1979; Clancy/Lee/Zoh 1986; De Villiers et al. 1979; Hakuta 1981; Sheldon 1974; Tavakolian 1981), psycholinguists observed that children (from 3 to 9 years) and adults performed similarly in relation to relative clauses comprehension. Furthermore, it was argued that the number of referents displayed in sentences with complex NPs could affect either cognitive processing cost or the capacity of children to analyze the sentences. Goodluck/Tavakolian (1982) claimed that difficulties in the comprehension of relative clauses were correlated to an increased number of [+animated] NPs since it would be cognitively demanding to keep three or more referents in one's active memory.

Corrêa (1995) questioned the fact that children from 3 to 6 years would perform similarly in relative clause comprehension, and proposed an alternative method to the standard acting-out procedure. An alternative methodology, it was argued, would be able to disentangle distinct processes that appear together in the standard acting-out procedure, specifically the process of relative clause analysis and the semantic/pragmatic comprehension of the content and action involved.

Corrêa (1995) developed some detailed criticisms of the standard acting-out procedure by listing 3 problems. First, the task instructions created a bias in the children's responses because it could lead them to treat center-embedded and right-branching relative clauses as independent sentences. This would happen not because the children had difficulties in analyzing the relative clauses, but because there were two distinct verbal actions. Second, the method did not allow for observing the particular demands of each type of relative clause, because the children's responses would be understood as a direct function of implicit grammatical knowledge. Third, the task

effectively mixed two types of activities, a contextualized game with toy animals that would represent actions in a decontextualized linguistic game, and sentences lacking communicative purposes. The acted-out sentences in the game lead the children to make use of semantic and pragmatic demands typical of drama, which gets confused with the objective of the task, which is to evaluate the capacity of participants to analyze and process relative clauses.

Corrêa (1995) observed that a task which only uses language without any commitment to real referents is most suitable for evaluating comprehension and for studying the acquisition of relative clauses. The task proposed by this author is an acting-out comprehension task that also allows the analysis of a child's mistakes. In this alternative task, children are told a story in which the animal characters do not behave, and the actions are expressed by relative clauses. Unlike the standard acting-out procedure, two animals of the same kind participate in each scene (for example, two horses, two cows, etc.). And unlike the standard acting-out procedure, in which each kind of animal is represented by only one toy, in the alternative task, each kind of animal is represented by two distinct characters in the same scene, as exemplified in (6a). Children are encouraged to pay attention to what the animals do in scenes such as (6a) in order to correctly select, for example, the horse that pushed the cow, when they are requested to act-out the animals' bad behavior, as in (6b).

(6) a. Um cavalo empurra uma vaca. Um outro cavalo caça a galinha.
'A horse pushes a cow. Another horse chases a hen.'

b. O cavalo que empurrou a vaca chuta a ovelha.
'The horse that pushed the cow kicks the sheep.'

Unlike the standard methodology, the instructions of the alternative task proposed by Corrêa (1995) do not inform the children of the number of actions, that is, they do not bias the number of actions to be acted out. Therefore, the way the children act out the sentences provides real information regarding their comprehension of the relative clauses. Since in the alternative task the children were only asked to act out the main clause, their answers could be interpreted without ambiguities, as a result of relative clause processing, which is the key to correctly selecting one of the two toy animals that are involved in the scene. In this way, the alternative task avoids the 3 problems discussed above.

The alternative task conducted by Corrêa (1995) aimed to observe whether there would be any changes in the linguistic development of relative clause acquisition by children from 3 to 6 years old, and how syntactic processing would be affected by the following independent variables: attachment (center-embedded or right-branching relative clauses), focus (with the relative marker working as the subject or the object of the relative clause), and animacy (animate or inanimate NPs). For example, (4) is a center-embedded relative clause that is subject-focused with [+animate] NPs.

The results reported in Corrêa (1995) contradicted previous findings from studies using the standard acting-out procedure, in that notable differences were found between children at 3 and 4 and between children at 4 and 5 years old. A main effect of attachment and focus was found. The number of correct answers significantly increased in children from 3 to 5 years old, stabilizing at 5. This age effect meant that the alternative procedure is more sensitive to successfully measure relative clause acquisition though the course of children's development.

The error rates in Corrêa (1995) were also considered as a dependent variable. Main effects of focus and animacy were found. The sentences that were most difficult to process, responsible for a greater number of mistakes, were the center-embedded ones, object-focused with 3 [+animated] NPs, as in (7).

(7) A galinha que o porco empurrou pula o carneiro.
'The chicken that the pig pushed jumps over the sheep.'

Finally, Corrêa (1995) described 4 steps of language development in children from 3 to 5 years based on relative clause comprehension skills: (i) identifying the relative marker (that/who) that introduced a predicate (noun modifier) expressed by a sentence (verbal category expansion); (ii) updating the canonical word order in order to recognize the relative marker; (iii) keeping the relative marker of the potential previous NP in working memory in order to eventually fill a gap (when there is an NP missing) in the relative clause; and (iv) keeping and initial complex NP in working memory until it could be analyzed in relation to the main clause verb.

3.2 Language processing studies

3.2.1 Coreference processing

Coreference processing is one of the most widely-studied sentence processing phenomena in Brazilian psycholinguistics. There are areas of psycholinguistic studies that look at binding processing in BP, testing anaphora, pronouns and referential expressions in several different syntactic positions and discursive contexts, manipulating different types of linguistic information and using different methods with different populations.

Although anaphoric resolution is mainly studied in the subject position, an innovative proposal was tested by Leitão (2005), who conducted five psycholinguistic experiments in BP, testing the role of animacy and syntactic parallelism in coordinate clauses with repeated names, overt and null pronouns in the direct object position, in sentences like (8).

(8) Os colegas pintaram Leo_i/o $quadro_j$ no camarim, mas depois esqueceram Leo_i/$ele_{i,j}$/$Ø_{i,j}$ no palco
'The colleagues painted Leo_i/the $paiting_j$ in the dressing room, but then they forgot Leo_i/$him_{i,j}$/$Ø_{i,j}$ on the stage.'

In the priming experiment, using a probe recognition task, 24 participants were instructed to read a sentence like (8) on a computer screen and decide whether a specific word did or did not appear in the sentence just read. In the experimental conditions, the probe words were the antecedents. Leitão (2005) showed that both overt and null pronouns were equally capable of having their antecedents retrieved, confirming his initial hypothesis. However, no animacy effects were found.

In the self-paced reading experiment, 22 participants were asked to read sentence segments on a computer screen and to press a button after reading each segment. The time it takes the participants to press the button is recorded in milliseconds so that processing can be measured and so as to check which conditions were read more quickly by the participants. The sentences were like (8), with the exception that the antecedents were names and the referring expressions could be repeated names or overt pronouns. Leitão (2005) found a penalty for anaphoric direct objects in BP, which, following the Repeated Name Penalty (Gordon/Grosz/Gillion 1993), and Centering Theory[2] (Grosz/Joshi/Weinstein 1995), means that rather than repeated names, even pronouns in the object position with antecedents also in non-salient positions, as example (8) shows, are the most psychologically efficient coreferential expressions.

In the second self-paced reading, Leitão (2005) compared two kinds of referring NPs—hyperonyms and hyponyms—in order to test which would more easily allow for the retrieval of inanimate antecedents. The results of an experiment with 26 participants confirmed the Informational Load Hypothesis,[3] that is, that hyperonyms rather than hyponyms allowed for the retrieval of antecedents more efficiently, since the latter contain a greater information, and this overloads working memory.

Leitão (2005) also conducted an offline experiment with 40 participants, who were asked to say whether a probe sentence was true or false after reading an ambiguous sentence like (9).

2 The Centering Theory, proposed by Grosz/Joshi/Weinstein (1995), is concerned with the relationship between coreference and coherence. According to the theory, some discursive entities are more central than others, which influences the speaker's use of coreferential expressions such as full NPs and pronouns.

3 The Informational Load Hypothesis was proposed by Almor (1999) and combines pragmatic factors relating to discourse and coreference. It predicts that there is an inverse correlation between the antecedent's discursive salience and the informational load carried by the anaphors. Thus, highly salient antecedents retrieved by highly informational anaphors overload working memory, causing processing difficulties.

(9) Paulo$_i$ entregou o Marcos$_j$/o revólver$_j$ para a polícia, mas depois retiraram ele$_{i/j}$/Ø$_{i,j}$ da investigação.
'Paul$_i$ handed over Marcus$_j$/the gun$_j$ to the police, but then they drew him$_{i/j}$/Ø$_{i,j}$ from the investigation.'

Probe sentence:
Retiraram o Marcos/o revólver/Paulo da investigação.
'They drew Marcus/the gun/Paul from the investigation.'

The results indicated that overt pronouns tend to retrieve animate antecedents in object positions more easily, like *Marcus* in (9). This is evidence in favor of syntactic parallelism since both antecedents and pronouns have the same syntactic position here. Furthermore, this finding corroborates the Informational Load Hypothesis, since overt pronouns seem to be better at retrieving non-salient antecedents, such as those in direct object positions. In his final experiment, a self-paced reading experiment with non-ambiguous sentences, Leitão (2005) replicated the results found in the offline experiment.

Taking the five experiments together, Leitão (2005) shed light on relevant issues concerning coreference processing in BP, suggesting that not only does it involve discursive aspects related to topichood, as highlighted by the Centering Theory, it also involves aspects such as animacy and syntactic parallelism, even in the resolution of the direct object anaphora.

In the past, BP was considered a pro-drop language, that is, a language which allowed null subjects. However, due to changes in the subject pronoun and verb morphology system, modern BP is now a semi-pro-drop language (↗6 Historical syntax). First, two novel subject pronouns were incorporated into the pronominal system: *você*, a singular second-person pronoun, and *a gente*, a plural first-person pronoun. Both *você* and *a gente* agree with singular third-person forms. Second, *você* and *a gente* compete with the old forms *tu*, a singular second-person pronoun, and *nós*, a plural first-person pronoun. However, *tu* and *nós* (especially *tu*) are also being used with singular third person agreement in non-standard BP varieties. We might also note that in non-standard varieties, *eles*, a plural third-person pronoun, is also being used with singular third person agreement. To sum up, the verbal paradigm in modern BP has reduced towards third person agreement; consequently, the use of overt pronouns is becoming more frequent.

Almor et al. (2017) examined coreference processing with null pronouns in BP. The authors replicated an experiment conducted in Spanish, which is a pro-drop language. On the one hand, they expected to find the Repeated Name Penalty; on the other, because the use of overt pronouns is becoming more frequent in modern BP, they did not expect to find the Overt Pronoun Penalty (Gelormini-Lezama/Almor 2011). The Overt Pronoun Penalty is a consequence of the Informational Load Hypothesis, and occurs when overt pronouns retrieve highly salient discursive antecedents, causing processing delays. Almor et al. (2017) conducted a self-paced reading experi-

ment in BP with 47 participants. The experimental sentences contained overt and null pronouns or repeated names retrieving antecedents either in the subject position (salient entity), as in (10a) below, or in the object position (non-salient entity), as in (10b). Rather than being exposed to segments, participants were exposed to entire sentences, one after the other on the computer screen; for example, one of the sentences in (10) was followed by one of the sentences in (11) according to each condition. The sentences were controlled for length, with the repeated name and overt pronoun conditions containing 15 syllables and the null pronoun condition containing 13.

(10) a. salient entity condition:
Carla desafiou Luís para uma partida de xadrez
'Carla challenged Luís to a chess match.'

b. non-salient entity condition:
Luís desafiou Carla para uma partida de xadrez.
'Luís challenged Carla to a chess match.'

(11) a. repeated name:
Carla o venceu rapidamente e sem esforço.
'Carla beat him quickly and easily.'

b. overt pronoun:
Ela o venceu rapidamente e sem esforço.
'She beat him quickly and easily.'

c. null pronoun:
Venceu-o rapidamente e sem esforço.
'NULL beat him quickly and easily.'

Against their expectations, the results showed the Overt Pronoun Penalty, just like in Spanish, which means that overt pronouns as in (11b) referring back to salient antecedents, as in (10a), cause processing costs despite the fact that overt pronouns are becoming very frequent in current BP usage. In addition, Almor et al. (2017) found a weak effect of the Repeated Name Penalty, since there was no significant difference between null pronouns, as in (11c), and repeated names, as in (11a), referring back to salient antecedents, as in (10a). In order to confirm these findings, the authors tested the same materials used in the self-paced reading experiment in an acceptability judgment experiment with 47 participants. The results indicated that speakers of BP found not-locally more acceptable overt pronouns, as in (11b), than null pronouns, as in (11c), retrieving salient antecedents in the subject position, as in (10a). Furthermore, the results indicated that null pronouns, as in (11c), were considered more acceptable than repeated names, as in (11a), for salient antecedents, as in (10a), which is in line

with the Repeated Name Penalty, which seems to be present across the languages of the world.

Almor et al. (2017) also conducted a corpus study to check whether overt pronouns are indeed becoming more frequent in the same kinds of verbal contexts they used in their experiments, in that they still seem to carry processing costs when referring to salient antecedents. Their conclusion is that null pronouns are still more frequent than overt pronouns despite the claims of linguists and the intuitions of speakers.

Carminati (2002) sought to finding support for the Position Antecedent Hypothesis, a parsing strategy for intrasentential anaphora which suggests that, as a sort of division of labor, the human sentence processor assigns more subject antecedent for null pronouns (NP) and more object for overt pronoun (OP). Subsequently, Teixeira/Fonseca/Soares (2014) investigated BP by running an eye tracking reading study with ambiguous intrasentential backward and forward anaphoras in complex sentences, these composed of subordinate clauses having either an overt or a null pronoun in the subject position. The participants clearly preferred to link the null pronouns to the subject position, regardless of the anaphora being backward or forward. However, the overt pronoun was only more frequently co-indexed with the object position in the backward anaphora, and without statistical significance.

It is noteworthy here that null pronouns are more frequent in academic and fiction writing than in the news and in spoken language. The conclusion is that BP's transition from a pro-drop to a non-drop language is an on-going process.

Besides the contributions on null pronouns, coreference studies in BP were also able to provide new evidence for binding processes as described in the literature. Psycholinguists have thus tested whether the generative binding principles are psychologically real. Chomsky (1993) proposed three universal binding principles: Principle A posits that reflexives and reciprocals must be bound to their antecedents in their local domain; Principle B postulates that pronouns must be free in their local domain, that is, they cannot be bound in their local domain; while Principle C states that referential expressions such as NPs must be free.

In English, most studies on binding processing have shown that Principle A and Principle B (especially the former) work as soon as coreference processing begins, acting like an initial filter blocking all antecedent candidates that violate structural constraints. The only cases of structurally unacceptable antecedent candidates that seem to be taken into account by the processor are those lacking an adequate structural acceptable antecedent in the sentence (Nicol/Swinney 1989; Clifton Jr./Frazier/Deevy 1997; Sturt 2003; Xiang/Dillon/Phillips 2009; Dillon et al. 2013; Chow/Lewis/Phillips 2014; Cunnings/Patterson/Felser 2015, among others). In those cases, structurally unacceptable antecedent candidates would be considered as potential antecedents, despite the fact that they violate the binding principles, because they would agree in gender/number with the referring expression. In this way it seems that, besides structural constraints, morphological agreement also plays an important role in coreference processing.

The issue is that English does not appear to be very suitable to test the influence of morphology on binding processing since it does not have redundant and visible morphological markings such as in Romance languages. BP, then, can fill a gap in the literature regarding this question.

Oliveira/Leitão/Araújo (2013) investigated how the reflexives *ele*/a *mesmo*/a 'himself/herself' were processed in a self-paced reading experiment with 24 participants. The sentences used contained more than one antecedent candidate that agreed in gender with the reflexives, one within their local domain (respecting Principle A), and the other outside their local domain (violating Principle A) like (12) and (13).

(12) Maria disse que Lilian machucou ela mesma no parque de diversão.
'Mary said that Lilian hurt herself in the amusement park.'
Lilian se machucou?
'Did Lilian get hurt?'

(13) João disse que José machucou ele mesmo no parque de diversão.
'John said that Joseph hurt himself in the amusement park.'
Jose se machucou?
'Did Joseph get hurt?'

The dependent variables were the gender of the structurally acceptable candidates and the gender of the structurally unacceptable candidates, plus the gender of the reflexives. The dependent variables were the reading times at the pronoun region and the region following the pronoun, known as the spillover region, and the answers to the comprehension questions. Oliveira/Leitão/Araújo (2013) did not find any morphological effects on gender. Also, the only antecedent candidates that were taken into account by the processor as potential antecedents were the structurally acceptable antecedents, in that no effects were found for the structurally unacceptable candidates. Leitão et al. (2014) found similar results when replicating the self-paced reading experiment using the same stimuli, (12) and (13), in an eye tracking reading experiment. However, when analyzing the offline results, the answers to the comprehension questions revealed that in sentences where there was a structurally acceptable feminine antecedent and a structurally unacceptable masculine antecedent that agreed with the reflexive, the participants answered the comprehension question at levels of chance, that is, they did not know the correct answer. Hence Oliveira/Leitão/Araújo (2013) concluded that Principle A is infallible in BP, meaning that it works as an initial filter blocking the influence of morphological cues coming from candidates that violate Principle A.

Leitão/Peixoto/Santos (2008) conducted two self-paced reading experiments in order to examine the role of gender, number, and animacy in Principle B processing in BP. In the first experiment, they tested whether an NP in the subject position could be bound to a pronoun in the object position, as in (14), which in theory would violate Principle B since they are in the same local domain. The independent variables were

the gender, number, and animacy features shared by the antecedent candidates and the pronouns, while the dependent variable was the reading time at the pronoun region and the spillover region.

(14) Tião/Talita/Os motoristas/As carretas atropelou(aram) ele imprudentemente na estrada de Cabedelo.
'Tião/Talita/The drivers/The trucks recklessly hit him on Cabedelo road.'

No effects of gender, number, or animacy were found, so structurally unacceptable antecedent candidates were not considered by the processor as potential antecedents. This is evidence in favor of the authors' initial hypothesis, that is, that Principle B acts like a filter blocking the influence of candidates that violate Principle B despite the fact they might agree with the pronouns. However, it should be noted that when singular masculine and structurally unacceptable animate candidates agreed with the pronouns, reading times at the spillover region were longer than when they did not agree.

In the second experiment, Leitão/Peixoto/Santos (2008) replicated the first experiment, with the exception that this time there was a structurally acceptable antecedent available in the discursive context, as in (15).

(15) Carlos atravessou a rua correndo. Tião/Talita/Os motoristas/As carretas atropelou(aram) ele imprudentemente na estrada de Cabedelo.
'Carlos crossed the street running. Tião/Talita/The drivers/The trucks recklessly hit him on Cabedelo road.'

Leitão/Peixoto/Santos (2008) replicated the results of the first experiment. The only difference was that reading times were faster in the second experiment than in the first. It was concluded that when there was no structurally acceptable antecedent candidate available, as in the first experiment, candidates that feature-matched the pronouns could be considered as potential antecedents even if they violated Principle B. However, when there was a structurally acceptable antecedent available, as in the second experiment, the search of an antecedent ended faster and the structurally unacceptable candidates were not considered. This is in line with the findings of Sturt (2003) for English.

Taken together, the results reported in Oliveira/Leitão/Araújo (2013) and Leitão/Peixoto/Santos (2008) corroborate the hypothesis of the psychological infallibility of the binding principles in BP. However, Alves (2019) conducted two eye-tracking experiments in BP manipulating the influence of structurally unacceptable antecedent candidates with two different gender features in BP, semantic and grammatical gender; semantic gender is related to the sex of the referent, while grammatical gender is arbitrary in the language. Unlike self-paced reading experiments used in Oliveira/Leitão/Araújo (2013) and Leitão/Peixoto/Santos (2008), eye-tracking experiments allow fine-grained measures related to the very beginning of linguistic processing. Eye-

trackers are able to record our eye movements throughout the text, calculating the fixation times, that is, how long our eyes stay on a certain part of the text, and the saccades, that is, the movements our eyes make going forward and backwards through the text. The reading measures collected by Alves (2019), the First Pass and First Fixation Duration, which correspond to very early reading processing stages, showed that structurally unacceptable antecedents can influence coreference processing despite the fact that they violate Principle B if they have semantic and masculine gender.

Alves (2019) reported that when both structurally acceptable antecedents and structurally unacceptable antecedents, called distractors, matched the pronouns, distractors with semantic gender caused facilitatory effects compared to grammatical gender, which might indicate antecedent mis-retrievals. This is evidence that semantic gender is more psychologically prominent in memory than grammatical gender. In addition, in sentences with mismatching structurally acceptable antecedents, matching masculine distractors caused longer reading times than feminine distractors, which is evidence that masculine gender is more prominent in memory than feminine. This latter effect has already been found previously, in Leitão/Peixoto/Santos (2008) for example. However, Leitão/Peixoto/Santos found it in the spillover region, at a late processing stage, and not at early processing stages, as in Alves (2019). This means that morphological cues influence pronominal antecedent retrievals much sooner than we thought, which leads to the conclusion that gender cues seem to be very relevant to pronoun antecedent retrievals in BP, and thus candidates that carry prominent gender cues can activate memory retrieval, leading to the early fallibility of Principle B. More studies on the processing of binding are needed in BP in order to clarify the role of morphology and structural constraints in the binding process.

3.2.2 Morphological processing

Attraction errors generally occur when verbs are inflected in the plural, despite their subjects being in the singular, because of the interference of an intervening head noun in the plural. In addition, there are semantic factors that can motivate error attractions such as a distributive or a collective interpretation of the subjects, syntactic factors such as the hierarchical position of the intervenient materials, or even discursive factors such as the linear distance between subjects and verbs. Attraction errors are failures of linguistic processing, and they can reveal how different kinds of information are combined in sentence building operations in real time, and can also reveal other issues that might compromise the autonomy of syntax. Furthermore, studies on attraction errors in BP are very fruitful due to the redundant and visible morphology of the language.

Corrêa/Rodrigues (2005) studied the production of attraction errors in the subject-verb agreement in BP. In their first experiment, they aimed to examine whether a long linear distance between the head subject and the verb, and the hierarchical position of

the intervening material, either a PP or a relative clause, would lead to more attraction errors. The authors expected that PPs rather than relative clauses would cause more attraction errors, since the former are structurally closer to the subject heads in the syntactic hierarchy. They asked 31 participants to listen to preambles followed by a verb in the infinitive that would appear visually on a computer screen. Participants were instructed to repeat the preamble and continue the sentence by inflecting the verb, as in (16).

(16) Short distance with an intervenient PP:
O diretor arrogante dos funcionários...
'The arrogant director of the employees...'

Short distance with an intervenient relative clause:
O jornalista que falou dos empresários...
'The journalist who has spoken of the businessmen...'

Long distance with an intervenient PP:
O instrutor calmo dos pilotos de avião...
'The calm instructor of the plane pilots...'

Long distance with an intervenient relative clause:
O ator que discordou dos críticos de teatro...
'The actor who disagreed with the theater critics...'

The results reported in Corrêa/Rodrigues (2005) showed that long linear distances between the subject head and the verb increased the number of attraction errors; however, no effects for the type of the intervening material were found. In their second experiment, the authors focused on determining whether the plural morphological marking of intervenient PPs would increase the number of attraction errors. They manipulated the number morphology of the head subject as well as the two following PPs that were modifying it. The experiment was conducted with 34 participants and the results showed that more attraction errors were produced when there were intervening PPs in the plural following a subject head in the singular. However, when the subject head was plural, the morphological marking of the intervening PP was not relevant, and far fewer attraction errors were found. Also of note is that that Corrêa/Rodrigues (2005) did not find any effects for the second intervening PP.

It was concluded that the hierarchical position of the intervening material is the most important factor that contributes to attraction errors, especially when the linear distance between the subject head and the verb is long, and when the subject head is singular followed by intervening material in the plural. The authors claimed that attraction errors do not necessarily compromise syntax autonomy and argue in favor of an incremental production model in which small units such as phrases could be sub-

mitted to morphophonological codification before the sentence is complete. They also argued that self-correction is evidence in favor of a monitoring processor that works simultaneously with the morphophonological codification and that is influenced by working memory limitations and accessibility issues of the grammatical codifications.

3.2.3 Syntactic processing of relative clauses

Relative clause processing in BP was investigated by Maia et al. (2007) in a study that aimed at presenting a clear dissociation between two processing stages—an early and a late one—which might explain the apparent contradictory results across languages with respect to the Garden-Path sentence processing model (cf. Frazier/Fodor 1978, Frazier 1979).

According to the Garden-Path model, which is based on principles of cognitive economy, the human sentence processor is an innate mechanism, and hence universal, and it regulates the incremental construction of syntactic structures in the simplest ways. The incremental construction of a syntactic structure is carried out by the human syntactic analyzer, which operates under two strategies, Minimal Attachment and Late Closure. These two principles regulate new phrases adding to previous syntactic structures by selecting, in first place, the simplest syntactic structure, by favoring adjunction to the last head, also known as the local head. Figure 1 illustrates the Garden-Path adjunction model with two representations of the phrase in (17):

(17) Arroz de polvo cozido
'Cooked octopus rice'

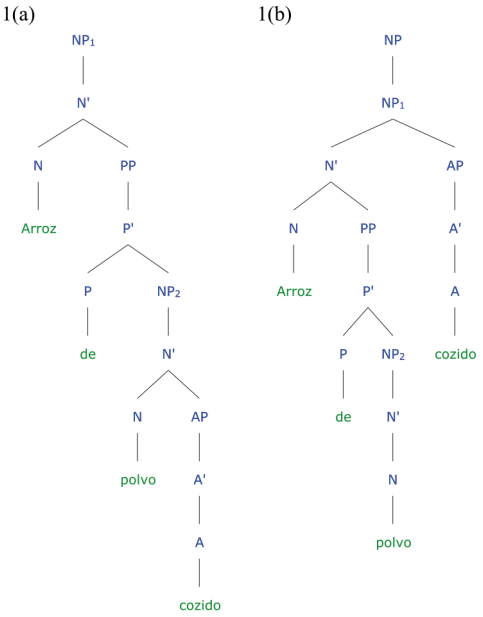


Figure 1: Simplified syntactic tree representations of local (1a) and non local (1b) attachment.

The representation in Figure 1(a) favors the principles of Minimal Attachment and Late Closure. Although (17) is permanently ambiguous, as the adjective *cozido* 'cooked' could be attached to either *polvo* 'octopus' or *arroz* 'rice', the parser prefers to attach it to NP2, *polvo* 'octopus'. If *cozido* 'cooked' was modifying the NP1, *arroz* 'rice', as shown in Figure 1(b), the structure would be more complex and cognitively more difficult to process.

If the adjective *pescado* 'fished' was used instead, the parser would not have the option of attaching the adjective to NP1 since only "octopus" could be fished, resolving the ambiguity. In English, unlike in Portuguese, 'Fished octopus rice', which corresponds to *Arroz de polvo pescado*, despite not being ungrammatical, is not semantically acceptable, and is not found in any corpus in English. The genitive structure in English favors Minimal Attachment and Late Closure, which does not occur in Portuguese.

Taking the difference between English and Portuguese into account, Maia et al. (2007) conducted a self-paced reading experiment which sought to monitor the early processing of relative clauses in BP. Previous data on Romance languages (Zagar/Pynte/Rativeau 1997; García-Orza et al. 2000; De Vincenzi/Job 1995; Cuetos/Mitchell 1988; Ribeiro 2005; Maia/Maia 2005) showed that speakers of French, Galician, Italian, Spanish, and Portuguese prefer NP1 attachment. In other words, the principles of Minimal Attachment and Late Closure are not universal, since these languages do not favor local attachment. However, Maia et al. (2007) argued that most of these studies

used offline experiments such as surveys to register the participants' conscious responses. In psycholinguistics, offline measures capture late processing stages after reanalysis. The responses given in a survey, for example, are influenced by discursive and pragmatic factors, as well as other factors related to gender and number agreement, which would interfere in the preference for local attachment.

Miyamoto (1999) investigated relative clauses in BP in a word-by-word moving-window non-cumulative self-paced reading experiment with bilingual speakers of BP and English living in the United States. Under examination were relative clauses with auxiliary verbs in either their inflected forms or in their participial forms, as shown in (18).

(18) a. O ator tentou ignorar a manchete das revistas [ø/que foi] **mencionada** no rádio.
b. O ator tentou ignorar as manchetes da revista [ø/que foi] **mencionada** no rádio.
'The actor tried to ignore the headline [sg/pl] of the magazines [sg/pl] [ø/that was] **mentioned** [sg] on the radio.'

The critical region analyzed by Miyamoto (1999) was the participial verb form, which is in bold in (18). The results indicated lower reading times in sentences like (18b), which suggests that local attachment is preferred in BP.

Since the results reported in Miyamoto (1999) contradicted previous findings in the literature, Maia/Maia (2005) conducted a survey comparing monolingual and bilingual speakers of BP and English in order to see whether the influence of L2 English could explain the findings in Miyamoto (1999) on the preference for local attachment in BP. Indeed, in Maia/Maia (2005), monolingual speakers of BP showed a preference for high attachment while bilinguals showed a preference for low attachment. Furthermore, Ribeiro (2005; 2008) also called into question the results of Miyamoto (1999) through a non-cumulative self-paced reading experiment, finding lower reading times in critical regions that favored high attachment, which corroborated Maia/Maia (2005) and other, previous studies on Romance languages.

Maia et al. (2007), in criticizing offline methods, attempts to solve the puzzle constituted by contradictory results on relative clause processing in BP. The authors conducted a self-paced reading experiment with 40 speakers of BP and 40 speakers of European Portuguese, with 24 experimental items and 48 fillers in a 2x2 design. The independent variables, besides the group, were the relative clause attachment (high or low) and the relative clause length (short or long). There was a comprehension question after the end of each sentence, as shown in the following example of an experimental item.

(19) a. A vítima reconheceu os cúmplices do ladrão que fugiram (depois do assalto ao banco).

b. A vítima reconheceu o cúmplice dos ladrões que fugiram (depois do assalto ao banco).
'The victim recognized the accomplice(s) of the thief (thieves) who ran away[3rdPl] (after the bank robbery).'

Question:
Quem fugiu?
'Who ran away?'

One of the hypotheses in Maia et al. (2007) is the Implicit Prosody Hypothesis (Fodor 1998), which is at the core of the argument of high attachment preference in surveys, since high attachment is preferred in long rather than short relative clauses (Finger/Zimmer 2005; Lourenço-Gomes 2003). Maia et al. (2007) also considered gender and number agreement as possible factors that would interfere in relative clauses resolution; however, none of these were in fact found to be relevant. The results in Maia et al. (2007) showed a main effect of the relative clause length, with Brazilian and Portuguese speakers taking an average of 1,807ms longer to read long relative clauses. Moreover, they found a main effect of relative clause attachment, as both groups needed an average of 281ms longer to read the items in the high attachment condition. With respect to the answers of the comprehension questions, an offline variable, the participants made more mistakes in the low attachment condition.

To sum up, the early preference of Portuguese speakers, either BP or EP, is for sentences with local attachment readings. Such a preference is a consequence of the locality principle of Late Closure. The results reported in Maia et al. (2007) can be reconciled with the results found in the survey study, which showed a high attachment preference. The authors explain that such a preference is not adopted by the parser. In other words, relative clause length interferes in attachment preferences only in late processing phases, possibly after the syntactic principles have been applied. Thus Maia et al. (2007) suggested that implicit prosody influences late attachment decisions, and it is considered to be a late-syntactic factor in sentence processing.

The conclusion is that the application of the Garden-Path principles—Minimal Attachment and Late Closure—is universal, as shown by Maia et al. (2007), where low attachment preference in early processing of relative clauses in BP was revealed. However, it is notable that at a late processing stage, high attachment can be preferred due to the Implicit Prosody Hypothesis, especially in long relative clauses. Therefore, Maia et al. (2007) showed evidence in favor of the universal nature of the Garden-Path model, showing that contradictory results can eventually be found in different materials, methodologies, and variables.

4 The emergence of psycholinguistics in the Brazilian Academy

The history of psycholinguistics in Brazil can be divided into 4 generations of researchers, who still work in the institutions where most psycholinguists in Brazil have emerged. We decided to call these universities "schools", and the term will be accompanied by the city where these institutions are located, for example, the Campinas school, the Rio school etc. We chose the expression "school" because it suggests a dimension of human resources training and education in a particular area of study, in research programs and experimental psycholinguistics laboratories. The expression 'psycholinguists generation' refers to the researchers who formed the second generation, and all of those who were formed by the first and by subsequent generations. The pioneers, the so-called first generation, are those who created or consolidated research programs and labs in Brazilian universities. The second generation refers to those who were supervised or academically aided by the pioneers, the third generation are those who were supervised by the second generation, and so on.

The first generation comprises those researchers who were interns or completed their doctorates in the United Kingdom or the United States. They were responsible for creating the first psycholinguistics labs in Brazil. It was in the second half of the 1970s, in the embryonic Campinas school, which refers to the city and the Universidade de Campinas (UNICAMP), in the state of São Paulo, where Cláudia Thereza Guimarães de Lemos began work following doctorate in 1975 at the University of Edinburgh under the supervision of John Lyons. De Lemos's career was not exclusively in psycholinguistics, since she worked in other fields of linguistics. She supervised students in the area of language and cognition. In 1976, she participated in the II Symposium on Cognitive Psychology, in the city of Gramado, in the Brazilian state of Rio Grande do Sul, where a study on the use of the Piagetian model in language acquisition research (De Lemos/Castro 1978) was presented. Eleonora Cavalcante Albano also began working at UNICAMP after her doctorate in the United States at Brown University with the dissertation *Phonological and lexical process in a generative grammar of Portuguese*. Albano (1987) founded a lab dedicated to speech studies and focused on phonetics, despite being a psycholinguist. Edson Françozo, also from the Campinas school, conducted pioneering work in his master's thesis in the late 1970s, investigating the electrophysiology of intonation. In his doctoral thesis, he studied the relationship between I-Language and aphasia, under the supervision of Marcelo Dascal. Having an interest in pragmatics and in the philosophical aspects of language, Françozo supervised doctoral dissertations on language processing in light of cognitivist and connectionist perspectives. At least 3 former Ph.D. students of Françozo founded research laboratories, at the Universidade Federal de Minas Gerais and the Universidade Federal do Rio Grande do Norte.

As part of the first generation, Leonor Scliar-Cabral founded the Santa Catarina school. Scliar-Cabral completed her doctorate at the University of São Paulo in 1977, and nowadays she is known as the pioneer of psycholinguistics in Brazil. In her doctoral dissertation she studied the acquisition of syntax and semantics in light of generative grammar. She dedicated her career to examining the underlying cognitive mechanisms of literacy. In her most-cited work (Morais et al. 1987), the authors found literacy effects in dichotic words processing. Many of the 56 master theses and 23 doctoral dissertations that Scliar-Cabral supervised were on reading and literacy. She also developed a literacy method based on psycholinguistics and neuroscience, and nowadays she works with teachers in the Northeast of Brazil. In the 1990s, Mailce Borges Mota arrived at the Santa Catarina school, founding the Laboratório da Linguagem e Processos Cognitivos, 'Laboratory of Language and Cognitive Processes'—LabLing, and working on language acquisition and L2 processing.

The first Brazilian university was founded in Rio de Janeiro, where several public universities are now located, and where two schools were developed. The Rio de Janeiro school is divided into two centers. The first of these is focused mainly on studying language acquisition, and the other is more focused on language processing. The Rio de Janeiro school was formed by the third pioneer of Brazilian psycholinguists, Letícia Maria Sicuro Corrêa. Her doctoral research was on a relative clause acquisition model (see section 3.1.3 in this chapter) and she defended the dissertation in 1986 at the University of London under the supervision of Jean Aitchison and John McShanne. Corrêa founded the Laboratório de Psicolinguística e Aquisição da Linguagem, 'Laboratory of Language Processing and Acquisition'—LAPAL, at the Pontifícia Universidade Católica do Rio de Janeiro (PUC Rio), which is dedicated to research on preferential looking, eye tracking, and other psycholinguistic methods. Several doctorate advisees of Corrêa have followed careers in psycholinguistics and have founded language acquisition labs in their own institutions, such as Maria Cristina Lobo Name, at the Universidade Federal de Juiz de Fora, who coordinates the Núcleo de Estudos em Aquisição da Linguagem e Psicolinguística, 'Center of Language Acquisition and Psycholinguistics'—NEALP.

The center of the Rio de Janeiro school that specializes in language processing is led by the fourth and last member of the first generation of Brazilian psycholinguists, Marcus Antonio Rezende Maia. Working at the Universidade Federal do Rio de Janeiro, Maia's PhD was completed at the University of Southern California in 1994 under the supervision of Maryellen MacDonald. His dissertation was on the anaphoric processing of the direct object in BP (see section 3.2.1 in this chapter). Maia founded the Laboratório de Psicolinguística Experimental, 'Laboratory of Experimental Psycholinguistics'—LAPEX, which conducts research on priming, eye tracking, and other experimental methods. Maia supervises master theses and doctoral dissertations. His former PhD students have created psycholinguistics laboratories in several different universities in Brazil, for example, Márcio Martins Leitão, who founded the Laboratório de Processamento Linguístico, 'Laboratory of Linguistic Processing'—LAPROL, at the Universidade Federal da Paraíba, in the Northeast of Brazil.

To sum up, the origins of psycholinguistics in Brazil rest on the shoulders of 4 researchers—De Lemos, Scliar-Cabral, Corrêa, and Maia—who comprise the first generation of psycholinguists. From the institutions where these pioneers work, research has developed and amplified over the years, reaching all Brazilian regions and various universities: Pontifícia Universidade Católica do Rio Grande do Sul, Universidade do Estado do Rio de Janeiro, Universidade Estadual do Ceará, Universidade Federal do Acre, Universidade Federal do Ceará, Universidade Federal Fluminense, Universidade Federal de Juiz de Fora, Universidade Federal de Minas Gerais, Universidade Federal da Paraíba and Universidade Federal do Rio Grande do Norte, among others.

5 Final remarks

This chapter has presented some of the most notable psycholinguistics studies conducted on Brazilian Portuguese. In the field of language acquisition, we reviewed works on the acquisition of gender and number agreement, lexical and functional morphology, and relative clauses, while in the field of language processing, we reviewed works on number morphology agreement, coreference, and relative clauses. We also discussed other important themes such as the semi-pro-drop status of BP, some of the differences between EP and BP, and some of differences between the standard and the non-standard varieties of BP. Furthermore, we showed how BP corroborates linguistic principles such as the Repeated Name Penalty and the Minimal Attachment and Late Closure universals. Finally, we showed how BP can answer fundamental questions in a number of the areas of psycholinguistic research, for example, the morphological agreement within the NP and in the subject-verb dependency, and the role of morphology in language acquisition and in coreference processing. As one can see, the studies carried out in BP shed light on important questions not only in the field of linguistics in relation to the organization of the lexicon and language structure, but also in the field of cognitive psychology, revealing, for instance, the role of the input and the task, memory organization and constraints, and how different kinds of information work together in our mind.

In this chapter we have introduced some of the methods and techniques used in psycholinguistics to collect data with different kinds of populations on a diversity of language phenomena, at different stages of linguistic processing, and with different kinds of purposes.

The field of psycholinguistics in Brazil has shown impressive development across the country, reaching a great number of universities and forming new generations of researchers.

Experimental research in psycholinguistics imposes strict methodologies in a multidisciplinary field that speaks to a number of related areas of study, such as artificial intelligence, neuroscience, biolinguistics, statistics, psychology, etc. There are a variety of themes that can be investigated from the perspective of psycholinguistics,

going beyond the boundaries of BP structural questions and in the direction of L1 and L2 acquisition models, bilingualism, literacy, etc.

6 References

Albano, Eleonora (1987), *O psicolinguista convertido*, Cadernos de Estudos Linguísticos 13, 41–48.

Almor, Amit (1999), *Noun-phrase anaphora and focus: The informational load hypothesis*, Psychological Review 106/4, 748–765, https://doi.org/10.1037/0033-295X.106.4.748.

Almor, Amit, et al. (2017), *Language processing, acceptability, and statistical distribution: a study of null and overt subjects in Brazilian Portuguese*, Journal of Memory and Language 92, 98–113, http://dx.doi.org/10.1016/j.jml.2016.06.001.

Alves, Michele Calil dos Santos (2019), *As diferenças entre gênero gramatical e gênero semântico na recuperação de antecedentes pronominais em português brasileiro*, Diacrítica 33/2, 89–115.

Bagetti, Tatiana (2009), *Um estudo experimental do processamento na interface fônica e da análise sintática inicial: o papel de elementos funcionais na aquisição da linguagem*, Doctoral thesis, Rio de Janeiro, Pontifícia Universidade Católica do Rio de Janeiro.

Bagetti, Tatiana/Corrêa, Letícia Maria Sicuro (2011), *The early recognition of verb affixes: evidence from Portuguese*, in: Nick Danis/Kate Mesh/Hyunsuk Sung (edd.), *BUCLD 35 Online Proceedings Supplement*, https://www.bu.edu/bucld/files/2011/07/BAGETTI-CORREA-BUCLD-35.pdf, last accessed 18.02.2022.

Bagetti, Tatiana/Corrêa, Letícia Maria Sicuro (2013), *O reconhecimento do padrão morfofonológico dos afixos verbais na aquisição inicial do português brasileiro*, Veredas 17, 20–38.

Berko, Jean (1958), *The child's learning of English morphology*, Word 14/2–3, 150–177, https://doi.org/10.1080/00437956.1958.11659661.

Bowerman, Melissa (1979), *The acquisition of complex sentences*, in: Paul Fletcher/Michael Garman (edd.), *Studies in language acquisition*, Cambridge, Cambridge University Press, 285–305.

Carminati, Maria Nella (2002), *The processing of Italian subject pronouns*, Doctoral thesis, Amherst (MA), University of Massachusetts.

Castro, Ana/Ferrari-Neto, José (2007), *Um estudo contrastivo do PE e do PB com relação à identificação de informação de número no DP*, Letras de hoje 42/1, 65–76.

Chomsky, Noam (1957), Syntactic Structures, The Hague/Paris, Mouton.

Chomsky, Noam (1993), *A minimalist program for linguistic theory*, in: Ken Hale/Samuel J. Keyser, *The view from building 20: essays in linguistics in honor of Sylvain Bromberger*, Cambridge, MA, MIT Press, 1–52.

Chomsky, Noam (1995), *The Minimalist Program for Linguistic Theory*, Cambridge, MA, MIT Press.

Chow, Wing-Yee/Lewis, Shevaun/Phillips, Colin (2014), *Immediate sensitivity to structural constraints in pronoun resolution*, Frontiers in Psychology 5, 630, https://doi.org/10.3389/fpsyg.2014.00630.

Clancy, Patricia M./Lee, Hyeonjin/Zoh, Myeong-Han (1986), *Processing strategies in the acquisition of relative clauses: Universal principles and language specific relations*, Cognition 24, 225–262, https://doi.org/10.1016/S0010-0277(86)80003-8.

Clifton Jr., Charles/Frazier, Lyn/Deevy, Patricia (1997), *Feature manipulation in sentence comprehension*, Rivista di Linguistica 11, 11–39.

Corrêa, Letícia Maria Sicuro (1995), *An alternative assessment of children's comprehension of relative clauses*, Journal of Psycholinguistic Research 24, 183–203, https://doi.org/10.1007/BF02145355.

Corrêa, Letícia Maria Sicuro/Almeida, Diogo/Porto, Renata Sobrino (2004), *On the representation of Portuguese gender-inflected words in the mental lexicon*, Brain and Language 90/1–3, 63–73, https://doi.org/10.1016/S0093-934X(03)00420-6.

Corrêa, Letícia Maria Sicuro/Augusto, Marina/Ferrari-Neto, José (2006), *The early processing of number agreement in the DP: evidence from the acquisition of BP*, in: David Bamman/Tatiana Magnitskaia/Colleen Zaller (edd.), *Online Proceedings Supplement of the BUCLD 30: Proceedings of the 30th Annual Boston University Conference on Language Development*, Somerville, MA, Cascadilla Press, http://www.bu.edu/bucld/files/2011/05/30-CorreaBUCLD2005.pdf, last accessed 18.02.2022.

Corrêa, Letícia Maria Sicuro/Name, Maria Cristina (2003), *The processing of Determiner-Noun agreement and the identification of the gender of Nouns in the early acquisition of Portuguese*, Journal of Portuguese Linguistics 2/1, 19–43, http://doi.org/10.5334/jpl.34.

Corrêa, Letícia Maria Sicuro/Rodrigues, Erica (2005), *Erros de atração no processamento da concordância sujeito-verbo e a questão da autonomia do formulador sintático*, in: Marcus Antonio Rezende Maia/Ingrid Finger (edd.), *Processamento da Linguagem*, Pelotas, Educat, 303–335.

Cuetos, Fernando/Mitchell, Don C. (1988), *Cross-linguistic differences in parsing: restrictions on the use of the Late Closure strategy in Spanish*, Cognition 30, 73–105, https://doi.org/10.1016/0010-0277(88)90004-2.

Cunnings, Ian/Patterson, Clare/Felser, Claudia (2015), *Structural constraints on pronoun binding and coreference: evidence from eye movements during reading*, Frontiers in Psychology 6, 840, https://doi.org/10.3389/fpsyg.2015.00840.

De Lemos, Cláudia T. G./Castro, Maria Fausta Pereira (1978), *Algumas observações sobre a utilização do modelo piagetiano em recentes estudos de aquisição da linguagem*, Cadernos de Estudos Lingüísticos 1, 51–63.

De Villiers, Jill G., et al. (1979), *Children's comprehension of relative clauses*, Journal of Psycholinguistics Research 8/5, 499–518, https://doi.org/10.1007/BF01067332.

De Vincenzi, Marcia/Job, Remo (1995), *An investigation of Late Closure: The role of syntax, thematic structure, and pragmatics in initial and final interpretation*, Journal of Experimental Psychology: Learning, Memory, and Cognition 21/5, 1303–1321, https://doi.org/10.1037/0278-7393.21.5.1303.

Dillon, Brian, et al. (2013), *Contrasting intrusion profiles for agreement and anaphora: Experimental and modeling evidence*, Journal of Memory and Language 69/2, 85–103, https://doi.org/10.1016/j.jml.2013.04.003.

Fantz, Robert L. (1961), *The origin of form perception*, Scientific American 204/5, 66–72, https://doi.org/10.1038/scientificamerican0561-66.

Finger, Ingrid/Zimmer, Márcia Cristina (2005), *A preferência de interpretação de orações relativas curtas e longas em português brasileiro*, in: Marcus Antonio Rezende Maia/Ingrid Finger (edd.), *Processamento da Linguagem*, Pelotas, Educat, 111–130.

Fodor, Janet Dean (1998), *Learning to parse*?, Journal of Psycholinguistic Research 27/2, 285–319, https://doi.org/10.1023/A:1023258301588.

Frazier, Lyn (1979), *On comprehending sentences: Syntactic parsing strategies*, Doctoral thesis, Storrs (CT), University of Connecticut.

Frazier, Lyn/Fodor, Janet Dean (1978), *The sausage machine: A new two-stage parsing model*, Cognition 6, 291–325, https://doi.org/10.1016/0010-0277(78)90002-1.

Galves, Charlotte (2001), *Ensaios sobre as gramáticas do português*, Campinas, Unicamp.

García-Orza, Javier, et al. (2000), *High attachment preferences in Galician relative clauses: Preliminary data*, Poster presented at AMLaP (Architectures and Mechanisms of Language Processing) 2000, Leiden, NL, September 20–23.

Gelormini-Lezama, Carlos/Almor, Amit (2011), *Repeated names, overt pronouns, and null pronouns in Spanish*, Language and Cognitive Processes 26/3, 437–454, https://doi.org/10.1080/01690965.2010.495234.

Goodluck, Helen/Tavakolian, Susan (1982), *Competence and processing in children's grammar of relative clauses*, Cognition 11, 1–27, https://doi.org/10.1016/0010-0277(82)90002-6.

Gordon, Peter C./Grosz, Barbara J./Gillion, Laura A. (1993), *Pronouns, names, and the centering of attention in discourse*, Cognitive Science 17, 311–347, https://doi.org/10.1207/s15516709cog1703_1.

Grosz, Barbara J./Joshi, Aravind K./Weinstein, Scott (1995), *Centering: A framework for modeling the local coherence of discourse*, Computational Linguistics 21/2, 203–225.

Hakuta, Kenji (1981), *Grammatical description versus configurational arrangement in language acquisition: The case of relative clauses in Japanese*, Cognition 9, 197–226, https://doi.org/10.1016/0010-0277(81)90001-9.

Harris, James W. (1991), *The Exponence of Gender in Spanish*, Linguistic Inquiry 22/1, 27–62.

Hauser, Marc D./Chomsky, Noam/Fitch, W. Tecumseh (2002), *The Faculty of Language: What Is It, Who Has It, and How Did It Evolve?*, Science 298/5598, 1569–1579, https://doi.org/10.1126/science.298.5598.1569.

Höhle, Barbara/Weissenborn, Jürgen (2000), *The origins of syntactic knowledge: recognition of determiners in one year old German children*, in: Catherine Howell/Sarah A. Fish/Thea Keith-Lucas (edd.), *Proceedings of the 24th Annual Boston Conference on Language Development—BUCLD 24*, Somerville, Cascadilla Press, 138–149.

Jusczyk, Peter W./Luce, Paul A./Charles-Luce, Jan (1994), *Infant's sensitivity to phonotactic patterns in the native language*, Journal of Memory and Language 33/5, 630–645, https://doi.org/10.1006/jmla.1994.1030.

Kato, Mary A. (2005), *Gramática do letrado: Questões para a teoria gramatical*, in: Maria A. Marques, et al. (edd.), *Ciências da linguagem: Trinta anos de investigação e ensino*, Braga, CEHUM, 131–145.

Leitão, Márcio Martins (2005), *O Processamento do objeto direto anafórico no Português Brasileiro*, Doctoral thesis, Rio de Janeiro, Universidade Federal do Rio de Janeiro.

Leitão, Márcio Martins/Peixoto, Priscilla Cristina de Carvalho/Santos, Susana Thais Pedroza (2008), *Processamento da co-referência intra-sentencial em português brasileiro*, Veredas 2, 50–61.

Leitão, Márcio Martins, et al. (2014), *Evidências experimentais do processamento da correferência e dos princípios de ligação em Português Brasileiro*, Revista da Abralin 13/1, 289–332.

Levelt, Willem J. M. (1992), *Psycholinguistics: an overview*, in: William Bright (ed.), *International encyclopedia of linguistics*, vol. 3, Oxford, Oxford University Press, 290–294.

Levelt, Willem J. M. (2013), *A history of psycholinguistics: The pre-Chomskyan era*, Oxford, Oxford University Press.

Lima, Eloisa de Oliveira/Gesualdi, Aline da Rocha/França, Aniela Improta (2009), *O estágio dois no desenvolvimento linguístico pré-fala: a percepção de consoantes*, Revista Linguíſtica 5, 81–109.

Lourenço-Gomes, Maria do Carmo (2003), *Efeito do comprimento do constituinte na interpretação final de orações relativas estruturalmente ambíguas: Um estudo baseado na "Hipótese da Prosódia Implícita"*, Master's thesis, Rio de Janeiro, Universidade Federal do Rio de Janeiro.

Maia, Marcus Antonio Rezende/Maia, Juliana (2005), *A compreensão de orações relativas por falantes monolíngues e bilíngues de Português e de Inglês*, in: Marcus Antonio Rezende Maia/Ingrid Finger (edd.), *Processamento da Linguagem*, Pelotas, Educat, 163–178.

Maia, Marcus Antonio Rezende, et al. (2007), *Early and late preferences in relative clause attachment in Portuguese and Spanish*, Journal of Portuguese Linguistics 6/1, 227–250, http://doi.org/10.5334/jpl.151.

Manning, Christopher D., et al. (2020), *Emergent linguistic structure in artificial neural networks trained by self-supervision*, Proceedings of the National Academy of Sciences 117/48, 30046–30054.

Mattoso Câmara Jr., Joaquim (1970), *Estrutura da Língua Portuguesa*, Petrópolis, Vozes.

Mitchell, Don C./Brysbaert, Marc (1998), *Challenges to recent theories of language differences in parsing: Evidence from Dutch*, in: Dieter Hillert (ed.), *Sentence processing: A Crosslinguistic Perspective*, San Diego, Academic Press, 313–335.

Mitchell, Don C./Cuetos, Fernando (1991), *The origins of parsing strategies*, in: C. Smith (ed), *Current issues in natural language processing*, Austin, Center for Cognitive Science, University of Austin, 1–12.

Miyamoto, Edson Tadashi (1999), *Relative clause processing in BP and Japanese*, Doctoral thesis, Cambridge (MA), MIT, Dept. of Brain and Cognitive Sciences.

Miyamoto, Edson Tadashi (2005), *Orações relativas ambíguas e a homogeneidade do processamento de sentenças*, in: Marcus Antonio Rezende Maia/Ingrid Finger (edd.), *Processamento da Linguagem*, Pelotas, Educat, 71–90.

Morais, José, et al. (1987), *The effects of literacy on the recognition of dichotic words*, The Quarterly Journal of Experimental Psychology 39A/3, 451–465, https://doi.org/10.1080/14640748708401798.

Name, Maria Cristina (2008), *Pistas prosódicas, sintáticas e semânticas facilitadoras da identificação dos elementos das categorias N e ADJ*, Revista Veredas 2, 1–15.

Name, Maria Cristina/Corrêa, Letícia Maria Sicuro (2003), *Delimitação Perceptual de uma Classe Correspondente à Categoria Funcional D: Evidências da Aquisição do Português,* Fórum Lingüístico 3/1, 55–88.

Name, Maria Cristina/Teixeira, Sabrina/Uchoa, Danielle (2015), *Thirteen-month-old Infants use phonologically strong function words in word segmentation and categorization*, Revista Linguíʃtica 11/1, 71–84, https://revistas.ufrj.br/index.php/rl/article/view/4622, last accessed 18.02.2022.

Nicol, Janet/Swinney, David A. (1989), *The role of structure in coreference assignment during sentence comprehension*, Journal of Psycholinguistic Research 18, 5–19, https://doi.org/10.1007/BF01069043.

Oliveira, Rosana Costa de/Leitão, Márcio Martins/Araújo, Elioenai Macena de (2013), *A influência dos antecedentes vinculados e não vinculados no processamento da anáfora "ele (a) mesmo (a)"*, Revista do GELNE 15 (Número Especial), 127–145.

Oliveira, Rosana Costa de/Leitão, Márcio Martins/Henrique, Judithe Genuíno (2012), *A Influência dos Antecedentes Vinculados e não vinculados no processamento da anáfora "a si mesmo(a)"*, Revista Linguíʃtica 8/2, 71–85.

Osgood, Charles E./Sebeok, Thomas A. (edd.) (1954), *Psycholinguistics: a survey of theory and research problems,* Bloomington, Indiana University Press.

Pires, Acrisio/Rothman, Jason (2009a), *Acquisition of Brazilian Portuguese in late childhood: implications for syntactic theory and language change*, in: Acrisio Pires/Jason Rothman (edd.), *Minimalist inquiries into child and adult language acquisition: case studies across Portuguese*, Berlin/New York, De Gruyter, 129–154.

Pires, Acrisio/Rothman, Jason (2009b), *Disentangling sources of incomplete acquisition: an explanation for competence divergence across heritage grammars*, International Journal of Bilingualism 13/2, 211–238.

Ribeiro, Antonio João C. (2005), *Late closure em parsing no português do Brasil*, in: Marcus Antonio Rezende Maia/Ingrid Finger Maia (edd.), *Processamento da Linguagem*, Pelotas, Educat, 51–69.

Ribeiro, Antonio João C. (2008), *A abordagem good-enough e o processamento de frases do português do Brasil,* Revista Veredas 12/2, 61–75.

Roberts, Ian/Kato, Mary A. (edd.) (1993), *Português brasileiro: uma viagem diacrônica (homenagem a Fernando Tarallo)*, Campinas, Unicamp.

Shady, Michele E. (1996), *Infants' sensitivity to function morphemes*, Doctoral thesis, Buffalo (NY), State University of New York at Buffalo.

Shafer, Valerie L., et al. (1998), *An electrophysiological study of infants' sensitivity to the sound patterns of English speech*, Journal of Speech, Language and Hearing Research 41/4, 874–886, https://doi.org/10.1044/jslhr.4104.874.

Sheldon, Amy (1974), *The role of parallel function in the acquisition of relative clauses in English*, Journal of Verbal Learning and Verbal Behavior 13, 271–281, https://doi.org/10.1016/S0022-5371(74)80064-2.

Shi, Rushen/Werker, Janet F./Morgan, James L. (1999), *Newborn infants' sensitivity to perceptual cues to lexical and grammatical words*, Cognition 72/2, B11–21, https://doi.org/10.1016/s0010-0277(99)00047-5.

Staub, Adrian/Rayner, Keith (2007), *Eye movements and on-line comprehension processes*, in: M. Gareth Gaskell (ed.), *The Oxford Handbook of Psycholinguistics*, Oxford, Oxford University Press, 327–342.

Sturt, Patrick (2003), *The time-course of the application of binding constraints in reference*, Journal of Memory and Language 48/3, 541–562, https://doi.org/10.1016/S0749-596X(02)00536-3.

Swinney, David A. (1979), *Lexical access during sentence comprehension: (re)consideration of context effects*, Journal of Verbal Learning and Verbal Behavior 18/6, 645–659, https://doi.org/10.1016/S0022-5371(79)90355-4.

Tanenhaus, Michael K., et al. (1995), *Integration of visual and linguistic information in spoken language comprehension,* Science 268/5217, 1631–1634, https://doi.org/10.1126/science.7777863.

Tavakolian, Susan L. (1981), *The conjoined-clause analysis of relative clauses*, in: Susan L. Tavakolian (ed.), *Language acquisition and linguistic theory*, Cambridge, MA, MIT Press, 167–187.

Teixeira, Elisangela Nogueira/Fonseca, Maria Cristina Micelli/Soares, Maria Elias (2014), *Resolução do pronome nulo em português brasileiro: evidência de movimentação ocular*, Revista Veredas 18/1, 281–301.

Xiang, Ming/Dillon, Brian/Phillips, Colin (2009), *Illusory licensing effects across dependency types: ERP evidence,* Brain and Language 108/1, 40–55, https://doi.org/10.1016/j.bandl.2008.10.002.

Zagar, Daniel/Pynte, Joe/Rativeau, Sylvie (1997), *Evidence for early-closure attachment on first-pass reading times in French*, The Quarterly Journal of Experimental Psychology 50A/2, 421–438, https://doi.org/10.1080/713755715.

Index

A
Abralin 37, 472
abstraction 2, 10, 134, 148, 149, 157, 169, 176, 214, 220, 230, 237, 238, 241, 242, 287, 290, 291, 340, 396, 461
Academy 28, 34, 440, 448, 495, 559, 561, 568, 569
accent (dialect) 113, 301
accent (orthographic). *See* orthography
accent (stress). *See* stress
acceptability 106, 582, 597–601, 604
accommodation 210, 283, 534
accomplishment 40, 258, 465
accusative 32, 104, 145, 157–61, 164, 181, 183, 226, 235, 403, 407
achievement 465
acoustic 361, 362
acquisition 64, 71, 86, 89, 91, 92, 96–98, 101, 107, 164, 219, 313, 314, 316, 442, 443, 494, 564, 580–82, 584, 586–94, 607–10
  early a. 590
acronym 119, 369, 394, 396, 397, 518, 540
address form 157–60, 170, 181, 227, 404, 405, 560
adjacency 168, 350, 360
adjective 25, 87, 137, 145–48, 372, 373, 376, 377, 379, 382, 441, 454, 474, 475, 493, 526–29, 531, 534, 540, 584, 586, 592, 604
adjunct 166, 244, 466, 603
adverb 162, 163, 369, 373, 374, 382, 471, 472, 526, 528, 529
adversative 138
affix 134, 135, 137, 138, 144, 148, 150, 162, 164, 170, 371–73, 375–77, 379, 380, 385–87, 389, 390, 392–95, 581, 584, 588, 590, 591
affixoids 385, 386, 393, 394
affricate 115, 117, 120, 121, 126, 128, 130, 238, 300, 320, 351
Africa 29, 30, 54, 59, 60, 64–74, 81, 86–108, 115, 121, 183, 201–8, 214, 238, 252–54, 261, 437, 438, 442, 447, 448, 494–497, 529, 530, 535, 537, 544, 548, 573, 575
Africanism 205–7
Africans 59, 63, 64, 69, 71, 72, 86, 87, 89, 92–96, 107, 493–99, 501, 529, 556
Afro-Bolivian 100
Afro-Brazilian 71, 100–103, 106, 107, 193, 208, 223, 290, 321, 438, 489
agentive 410
agglutination 134, 138, 198, 518, 525, 526
agrammatical. *See* ungrammaticality
agreement 87–92, 97–107, 135, 136, 146, 147, 157, 170, 171, 179, 180, 215, 221, 223–28, 285, 292–95, 299, 301, 313, 320, 372, 373, 382, 404–407, 414, 421, 476, 566, 568–70, 572, 574, 579, 581, 584, 586–88, 590, 596, 598, 600, 601, 605, 606, 609
Aktionsart 466
Alagoas 255, 256, 261
Alencar, José de 29, 30, 232, 236, 245, 501, 557–60
Algarve 113, 122
alloglossia 554
allomorph 138, 143, 377, 378, 380
allophone 114, 116, 117, 126, 128, 129, 350
alphabetization 6
Amapá 272, 353
Amaral, Amadeu 31, 33, 136, 252–54, 256, 262, 266, 269, 284, 285
Amazonas 24, 26, 64, 67, 68, 70, 72, 81, 93, 113, 125, 271, 272, 342, 489, 490, 542, 555
ambiguity 170, 172, 176, 180, 406, 429, 444, 465, 593, 595, 598, 604
Ambundo 205
America 25, 35, 59, 67, 69, 89, 94–99, 122, 210, 286, 315, 316, 460, 490, 494, 516, 545, 557, 558, 563, 572
Amerindian 10, 27, 35, 86–89, 93, 95, 96, 107, 108
analogy 26, 139, 142, 143, 148, 391
analytical passive. *See* passive
anaphora 164, 165, 186, 428, 594–96, 598
Anchieta, José de 23, 25, 56, 490–92
Andalusian 122
Ando-Peruvian 518
Andrade, Mário 288, 560
Andrade, Oswald 520
Anglicism 211, 545
Angola 65, 69, 214, 497, 569–71
animacy 296, 584, 592–96, 599, 600
Anobon 106
antecedent 165, 168, 169, 244, 296, 595–601
anthropology 35, 212, 230, 524, 543

https://doi.org/10.1515/9783110405958-022

anthroponym 204, 513, 514, 531, 533, 538, 540–44, 546–48
anthropophagism 288, 520
applied linguistics 14, 197, 443, 445
Arabic 202, 204, 205, 438, 517, 541
Arawak 54, 93, 525, 526
archaic Portuguese. *See* Old Portuguese
archiphoneme 348, 349, 360
architecture 3, 193, 194
Argentina 309, 311, 317, 318, 324, 325
argument 11, 45, 155, 157, 162, 165, 182, 221, 224, 292, 295, 406–12, 416–23, 425, 429, 461–65, 473, 477
  biargumental 175, 224
article 10, 11, 32, 87, 102, 103, 146, 228, 324, 359, 534, 535
Asia 107, 310, 575
aspect 228, 373, 375, 465, 466, 471, 473
  progressive a. 228, 471
assimilation 10, 235, 344, 348, 349, 351
  regressive a. 351
Atlantic 106
atlas 34, 126, 196, 213, 251, 253, 256–73, 277, 343, 359, 361, 364
  multi-dimensional a. 251, 261, 269, 270, 273, 278
  one-dimensional a. 259, 261, 269, 278
  sound a. 272
  topodynamic a. 278
  topostatic a. 278
  two-dimensional a. 259–61, 269
augmentative 383, 536, 537
Aurélio 372, 373, 447, 448, 463
ausbau 488
auxiliary 162, 171–73, 222, 317, 416, 420–24, 471, 473, 605
Aymara 317
Azores 64, 116, 266

B

Bacongo 205
Bahia 61–65, 68, 69, 91, 100, 115, 119–21, 125, 160, 161, 166, 196, 197, 227, 259, 290, 297, 342, 438, 498, 522, 530, 534
Bakhtin, Michail 15, 468
bandeirantes 93, 489
Bantu 59, 65, 72, 81, 106, 149, 183, 201, 204, 205, 225, 227, 370, 438, 530
Barbosa, Rui 30, 222, 235, 236, 263, 501
bare nominal. *See* noun
behavioral methods 580, 581, 590
bestialogics 500
biargumental. *See* argument
Bible 491, 507, 508, 519
bilingualism 309–14, 317, 324, 326, 351, 357, 489, 554, 572, 580, 605
  near-b. 563
binding 169, 594
  b. principles 598–601
bi-nomination 519
biolinguistics 580
bleaching 158, 461, 546
blending 369, 391, 392, 394, 396, 397
Boléo, Manuel de Paiva 237, 238
Bolivia 96, 118, 317, 318
borrowing 139, 193, 201, 310, 393, 438, 439, 448, 460, 470
bottom-up change 75, 508, 509
Brasiguaios 325
Brasildeutsch 311
Brasília 3, 79, 270, 286, 289, 572, 574
Brazilian-American 325
Brazilianism 229–31, 237, 256, 447, 448, 452
brazilianization 501, 559
Brazilian-Japanese 314
Brazilianness 561

C

caipira 253, 254, 256, 269, 285, 301, 356
Calvinist 24
candomblé 438, 530
captaincies 57–61, 63, 64
Caribbean 86, 91, 94–96
carioca 3, 33, 253, 254, 256, 285, 342
cartography 179, 513, 516, 518, 542
catechism 26, 93, 203, 488, 490–93, 501, 509, 519, 553
catholic 488, 492, 499, 500, 508, 518–21, 534, 570
Ceará 120, 121, 261
Censo Project 6, 7, 297, 298
chain of varieties 285, 286
chiamento 123–25, 357
Chinese 309, 310, 314
christian 59, 61, 490–92, 496, 557, 571
church 500, 518–20
cleft 412–17, 429, 431
  pseudo-c. 416

clipping 392–97
clitic 10, 104, 157, 159–62, 164–66, 181, 186, 221–23, 226–29, 244, 295, 296, 352, 412, 420
coda 260, 261, 268, 274, 278, 301, 302, 312, 317, 318, 337, 338, 340, 341, 347, 349, 350, 354–60, 384
  word-internal c. 356–59
code-switching 287, 310, 316–18, 325
cognates 140, 318, 461, 475
cognition 289, 290, 296, 298, 374, 378, 379, 401, 460, 469, 480, 580, 592, 603–9
Coimbra 2, 56, 61, 554
co-indexing 416, 598
colloquial 114, 120, 122, 126, 133, 136, 137, 149, 181, 186, 207, 211, 375, 505, 565, 588–90
Colombia 67, 93, 96, 317
colonialism 69, 287
  pre-colonial 514
colonies 56, 64, 130, 234, 311, 494
colonization 3, 18, 27, 57, 59–77, 85–98, 107, 234, 288, 323, 437, 488–94, 508, 509, 514, 519–525, 528, 532, 533, 539, 553–57, 561, 572
communication 138, 139, 184, 185, 287, 409, 413, 416, 418, 420, 425, 426, 436, 439, 444, 445, 453, 454, 488, 491–509, 564, 573, 593
comparative 31, 34, 87, 133, 136, 137, 221, 224, 244, 431
competence 242, 245, 287, 288, 298, 316, 460, 472, 547, 589, 590
complex sentence. *See* sentence
complex system theory 41, 44, 45, 460
complex word. *See* word
composition (morphology) 142, 428, 519, 531, 538–40, 546
compositionality 140–42, 371, 375, 377, 378, 388, 390, 459, 466, 471
  non-compositional 141, 375
compound 377–80, 385–96, 444, 453, 461, 465, 471, 473, 476, 534, 536
  re-compounding 370, 391, 393, 394
computational 196, 268, 470, 472, 580
conceptual 499, 502, 507
Congo 65, 69
Congo-Cordofanian 59, 204
conjugation 117, 138, 145, 147, 210, 170, 376–78, 473
conjunction 138, 163, 413, 424, 426, 428, 461, 467, 476, 499
consonant 114–30, 149, 200, 275, 320, 321, 338–40, 346–54, 356–58, 360, 544, 567, 568, 570
  preconsonantal 121–23, 125
constituent 340, 363, 376, 387, 389, 392, 394, 406, 407, 409–14, 416, 417, 419, 429
constitution 78, 237, 444, 560–63, 568
constraint 31, 156, 285, 286, 290, 317, 339, 375, 382, 386, 390, 409, 423, 598, 601, 609
construction 106, 171–77, 180–87, 224, 228, 242, 244, 245, 377, 378, 387, 390–92, 394, 407, 409, 410, 412–29, 431, 461, 469
  ergative c. 413
contact 64–92, 96–108, 201–11, 309–14, 316, 351, 352, 357, 470, 479, 518
contrastivity 168, 169, 226, 355, 414, 416, 418, 425, 478
coordination 387, 389, 499
copula 88, 99, 105, 185, 413, 475
copying 395, 428–30
C-Oral-Brasil 5
C-Oral-Rom 5, 19
coreference 409, 411, 418, 428, 579, 581, 594–96, 598, 601, 609
coronal 337, 341, 343, 345, 351–53, 357–59
corpus 4–9, 13, 37–43, 155, 163–69, 173, 179, 180, 186, 193–98, 202, 215, 228, 297, 440, 445, 449, 450, 452, 463, 464, 472
Coseriu, Eugenio 18, 35, 220, 286, 487
CPLP 215, 439, 440, 449, 553, 571–73, 575
creole 30, 64, 68, 72, 76, 81, 86–89, 91, 92, 94–107, 156, 223, 227, 230, 234, 235, 314, 315
critical discourse analysis. *See* discourse
Cuiabá 525, 534
cuiabano 270
culture 30, 150, 285, 287, 381, 448, 470, 474, 488, 494, 496, 507, 508, 515, 520, 521, 525, 541–6, 557, 558, 571, 572
CV-syllable. *See* syllable
cycle 171, 339

D

dative 11, 19, 106, 157–59, 161, 181–83, 226, 227, 403, 407
deadjectival 148
deadverbial 148

decreolization 88, 89, 91, 98, 156
defectiveness 374
definite 10, 146, 179, 221, 227, 228, 241, 359, 411, 534–36
delateralized 120, 320
deletion 164, 184, 210, 337, 338, 352, 354, 355, 357, 359, 360, 384, 389
demography 53, 63, 65, 71–73, 76, 79, 80, 95, 98, 107, 205, 214, 283, 286, 315
demonstrative. *See* pronoun
denasalization 337, 360, 361
denominal 148
dental 129, 203, 357
deontic 475, 503
dephonologization 119
depronominalization 428
derivation (morphology) 39, 133–49, 350, 369–75, 379, 380, 382, 384–88, 390, 391, 393, 394, 397, 441, 475, 537, 548
  popular regressive d. 537
detelicization 471
determiner 102, 176, 227, 228, 234, 293, 320, 407, 584–88, 590
  pseudo-d. 585
diachrony 43, 116, 125, 126, 133, 135–37, 139–42, 144, 145, 147, 148, 150, 155, 156, 159, 165–68, 174, 178, 182, 183, 479
diagenerational 268, 272
dialect 113, 117, 120–124, 137, 141, 170, 183, 200, 214, 224, 233, 236, 251–57, 262–73, 275–289, 311, 312, 316–23, 341, 343, 358, 359, 523, 524, 530, 564
  micro-dialectal 309, 326
  subdialect 260, 268, 269
dialectology 122, 251–53, 257, 258, 260, 262–65, 277, 515
dialectometry 267
diaphasic 285, 488
diasexual 268, 272
diaspora 315, 325, 574, 575
diastratic 268, 273, 277, 285, 488
diasystem 2, 3
diatopic 251, 261, 269, 273, 275, 277, 278, 285
diatopic-cinetic 272
diatopic-topostatic 272
dictionary 15, 141, 194–96, 199, 200–12, 251, 440, 442, 443, 445–53, 472, 557, 569, 570
diglossia 93, 107, 493, 494, 501
diminutive 142, 145, 235, 312, 340, 379–82, 384, 385
diphthong 115–19, 127, 128, 130, 301, 312, 313, 348–50, 360
discourse 169, 170, 227, 244, 295, 376, 377, 381, 411, 414, 418, 419, 425, 472, 479, 594, 596, 600, 601, 605
  critical d. analysis 15
  d. analysis 15, 230, 234, 237, 460, 493, 494, 496, 498–500, 507, 556
  d. traditions 7, 14, 15, 292, 454, 455, 464, 479, 487–98, 502, 504–9
  universes of discourse 487, 493, 499
distance (communicative) 488, 491, 492, 499, 500, 503, 507–9
divergence 116, 126, 140, 157, 160, 194, 200, 228, 235, 380
diversity 2, 3, 29, 214, 491, 514, 561, 563, 572
drift 42, 85, 86, 88–90, 92, 99, 200, 225
Dutch 65, 114, 311, 522

E

early-acquired. *See* acquisition
East Timor 571, 573
Ecuador 316
educated norm/speech 2, 6, 15, 24–30, 35, 37, 40, 54, 61, 70–78, 86, 100–8, 114, 126, 211, 222, 223, 229, 234, 235, 239, 261–67, 273, 277, 278, 286–96, 320–24, 352, 356–62, 370, 371, 405, 443, 463, 498, 500–9, 544–47, 557, 560–65, 571–76, 590, 607
education 7, 24–30, 35, 36, 40, 53, 54, 61, 70–78, 100, 107, 108, 214, 219, 220, 286, 287, 320, 443, 500, 504, 557, 560–65, 571–74
effeminacy 301
electroencephalography 581
electrophysiology 607
elite 61, 86, 93, 99, 104–7, 231, 285, 493–95, 499–501, 504, 553
emphasis 117, 137, 168, 363
enclisis 162–64, 222, 223, 228, 234, 235
endocentric 388
English 142, 149, 209–11, 316, 325, 393, 438, 545
epenthesis 114, 115, 119, 121, 130, 354, 382–84, 499
epistemology 44, 194, 199

ergative construction. *See* construction
Ethiopian 65, 496
ethnicity 24, 30, 54, 62, 67, 68, 72, 76, 87, 88, 94, 95, 107, 255, 309, 352, 441, 442, 448, 522, 525, 530, 535, 541–44
ethnography 28, 253, 265, 284, 299, 300, 303, 326, 515, 524
ethno-linguistic 65, 205, 532
ethnology 26, 66, 204, 287, 515
ethnonym 540–42, 548
etymology 117, 134–42, 175, 193–96, 199–213, 393, 449, 514, 517, 525, 531, 532, 535, 537, 546, 548, 566
  folk e. 537
  pseudo-etymological 567
eurocentrism 12, 18, 44
Europe 25–28, 35, 59, 61, 64, 69, 75–77, 95, 96, 107, 135, 221, 229, 234, 277, 285, 436, 491, 492, 503, 514, 519, 520, 531, 557, 572, 573
European Portuguese 90, 113–39, 143, 150, 155–66, 171, 174–85, 219–28, 238–45, 344, 358, 363, 404, 447, 466, 588, 589, 605
event (semantics) 422, 465, 466, 471, 503
event-related potentials 581
evolution 63, 89, 113–30, 155–57, 164–69, 173, 178–81, 186, 187, 193, 225
evolutionism 220, 221, 229, 230, 236, 239
exclusivity 414, 416
exhaustivity 416
existential 174–79, 187, 228, 413
exocentric 388, 390
exonyms 516
exotic 12, 232, 542, 546–48
experiment 7, 469, 579–600, 605–9
expletive 180
external factors 13, 14, 16, 17, 53, 60, 88, 113
extralinguistic 213, 293, 543, 547, 548
extrametricality 340, 341
eye-tracking 600

F

face-saving 381
Fa d'Ambu 106
feature (grammatical) 87, 114–24, 128, 171, 223, 224, 284–87, 291, 292, 296, 301, 320, 321, 341, 344, 349, 358, 383, 396, 402, 420, 423, 584, 586, 600
female person 76–78, 103, 114, 119, 144–47, 213, 293, 299, 301, 360, 375, 543, 545, 554
feminine (gender). *See* gender
first-person 161, 187, 299, 318, 321, 322, 325, 377–79, 596
floating nasal. *See* nasal
focus 163, 224, 226, 236, 363, 412–16, 425, 593, 594
folk etymology. *See* etymology
foreignisms 209, 210, 235, 354, 438, 439, 441, 443, 451, 531, 538, 544, 545, 566
foreign language 215, 442, 443, 560, 574
foreignness 209
formal linguistics 39, 296, 297, 401, 402, 406, 409, 410, 414–17, 420–25, 431, 460, 466, 467, 470–72, 480
formal speech 37, 126, 137, 158, 166, 224, 287, 352, 412, 427, 488, 505, 564, 589
Franciscan 24, 521
Frege, Friedrich Ludwig Gottlob 467, 471
French 24, 63, 93, 139–42, 169, 209–11, 438, 521, 531, 548, 565
fricative 114, 116, 119–22, 128–30, 337, 347, 351–59
Fronterizo 319, 322–25
fronting 185
functionalism 39, 290, 296, 298, 401, 402, 406, 409, 410, 413, 416–23, 426, 431, 461
functions of grammar 104, 138, 140, 148, 157, 169, 171, 181, 182, 213, 219, 295, 337, 381, 388–90, 401, 413, 414, 417, 420, 422, 423, 425, 466, 474, 475, 503, 537, 588, 590, 591
fusion 159, 371, 526, 546

G

Galician 150, 310, 316, 358, 604
Gallicisms 565, 566
garden-path 603, 606
gender 103–5, 134, 136, 144–48, 318, 372–75, 384, 543, 584–86, 599–601
  feminine g. 103, 104, 134, 136, 140, 144–48, 167, 338, 372–75, 382–84, 392, 474, 536, 584–86, 599, 601
  masculine g. 144–47, 301, 374, 382, 543, 584–86, 599–601
  neuter g. 140
genderless 384
generative 134, 171, 289–91, 296, 297, 401, 402, 406, 409–11, 414, 418, 419, 429, 431, 460, 462, 598

generic 166, 228, 244, 376, 471, 478
genitive 157–61, 179, 224, 374, 403, 604
genre (textual) 15, 186, 454, 462, 491, 496, 501–9, 564
geography 73, 251, 258, 266, 271, 273, 283, 345, 532, 534
geolinguistics 251, 256–58, 261–65, 269, 270, 273, 277
geo-sociolinguistics 261, 262, 269–73, 277
German 76, 310, 311, 351, 357, 439, 442, 531, 532
Germanic 106, 149, 202, 310, 311, 538
gerund 13, 228, 235, 372, 424, 465, 473–77
gerundism 477
glide 300, 338, 349–54
glossonym 540–42, 548
glottal 121–23, 128–30, 274, 356
glottocide 80
Goiás 73, 256, 257, 264, 270, 271, 342
grammar (system) 31, 32, 39–42, 168, 169, 232, 233, 236, 239–42, 245, 402, 534
  mixed g. 30, 134, 136, 144, 159, 160, 317, 548, 560, 572
  subsystem 198, 207, 287
grammar (work) 25, 27, 40–42, 56, 61, 65, 67, 77, 113, 491, 492, 497, 565, 569
grammarians 32, 42, 221, 229, 235
grammaticality 582
grammaticalization 45, 98, 99, 106, 136, 159, 171, 172, 187, 294, 297, 424, 425, 460, 461, 479
Grão-Pará 64, 67, 70, 72
Greek 140, 141, 389
Guarani 325, 489, 515, 518, 525, 541, 543, 557
Guinea 59, 106, 571
Guinea-Bissau 103, 214, 570, 571

H

Haitian 94, 315
Hausa 204, 205
Helvécia 76, 89, 91, 100, 103, 321
heritage language 310–16, 325, 326, 588, 589
hiatus 119, 384, 567, 569
historical-comparative 32, 229, 239
historiography of linguistics 14, 23, 67, 219
history 1, 2, 97, 105, 115, 120, 136, 146, 148, 209, 402, 450, 513, 523, 541, 543
hodonym 532, 533, 541
homographs 569
homonym 134, 140, 141, 199, 200
Hong Kong 314
Houaiss, Antônio 203, 239, 248, 449, 451
Hunsrückisch 309–11
hybrid 159, 309, 316, 317, 323, 325, 393, 531, 535, 537
hydronym 441, 525, 541
hyperonym 141–48, 595
hyper-raising 180
hyphen 452, 570
hyponym 595

I

IBGE 208, 438, 513, 521, 532, 541, 543
ideology (linguistic) 1, 18, 41, 220, 240, 245, 518, 520, 548, 566
idiom 377, 378, 453
I-language 241–45, 607
illiteracy 77–80, 89, 95, 493, 496, 499, 502, 503
immediacy (communicative) 488, 491–94, 498–500, 503, 509
immigration 75–77, 81, 107, 122, 208–10, 309–19, 323–26, 561, 563
imperfect 321, 340, 349, 465, 503
imperfective 321, 376, 471, 482
impersonal 11, 420
implicature 467
implosive 115–21, 124–30
inclusive language 144
indefinite 137, 147, 179, 243
independence 64, 74, 76, 206, 207, 210, 231, 237, 557, 558, 560
indeterminacy 105, 226, 405, 421
indexicality 144–48, 283, 284, 300, 301, 395
Indian 56, 93, 107, 490–94, 554, 555
indigenous 27, 35, 36, 54, 56, 61, 62, 66–73, 80, 81, 85, 86, 92–94, 115, 201–4, 230, 232, 315, 448, 489–95, 563
Indo-European 134, 137, 242
industrialization 75–79, 88, 108
infinitive 122, 143, 172, 173, 228, 243, 285, 311, 320, 355, 424, 477
  inflected infinitive 407, 588–90
inflection 100–5, 164, 169–71, 179, 339, 341, 370–80, 387, 402–6, 588–90
informal speech 119, 130, 158, 207, 285, 287, 386, 405, 412, 424

innovation 115, 118, 123, 128, 130, 136–40, 157, 162, 181–87, 212, 220–23, 226–28, 240, 404, 424, 478
integration 27, 67, 95, 96, 215, 223, 290, 314, 452, 555, 563, 572
intensification 137, 257, 395, 529
interface 38, 171, 409, 418, 419, 581, 591
interference 115, 117, 314, 318
interfix 142, 143
interjection 376–79, 382
internal language 86–90, 118, 241, 242, 582
interrogative 169, 185, 186, 224, 244, 317, 361–64, 430, 431
  rhetorical interrogative 362, 363
interview 38, 39, 273, 297, 300, 507
intervocalic 121, 200, 318, 355, 567
intonation 130, 164, 337, 361–64, 607
intuition 41, 242, 373, 465, 475, 480, 598
inversion 179, 296
irregularity 134, 135, 147, 160
irregular linguistic transmission 85–87, 92, 97, 98, 101, 105–7, 213, 494
isogloss 260, 267, 278, 323
isolexic 263
isomorphism 387, 388
isophonic 263
Italian 309–12, 346, 351, 352, 357, 531, 541
Italic 140

J
Jamaican 94
Japanese 86, 146, 149, 310, 313, 325, 523
jargon 455
Jesuit 24–26, 60, 61, 65, 67, 70, 72, 93, 488, 491–96, 509, 553–55
Jews 59
journalism 500–5, 509
journalist 505, 561, 565
junction 499
juxtaposition 317, 424

K
Karib 93, 525, 526, 536
Karitiana 478
Khoisan 59
Kikongo 204, 205
Kimbundu 65, 204, 205, 227, 497, 530, 535
Kiriri 25
koine 113, 122, 311
Kordofanian 59
Korean 314

L
labiodental 129
Labov, William 283–90, 294–99, 302
language shift 89, 313, 317, 322
language teaching 14, 24, 28, 41, 42, 121, 215, 229, 234, 239, 494, 557, 563–65, 574, 575
lateral 129, 320, 351, 353, 356, 357
Latin 137–45, 148, 171, 200, 213, 461, 475, 476
learned form 114, 119, 130, 393, 589
lemma 203, 215, 451–53
letter 43, 44, 114, 119, 158, 160–63, 184–86, 488, 498, 502, 560
lexicalization 318, 324, 349, 371, 375, 378, 379, 381, 392, 453
lexicography 14, 141, 196–200, 203, 439–53
lexicology 134, 135, 138, 141, 144, 145, 440, 441, 445
lexicon 28, 134, 138–50, 193–215, 261, 263, 276–78, 379, 435–53, 473, 474, 580, 590, 591, 609
liaison 352
Libras 15, 81, 563
licensing 169, 177, 182, 227, 410, 411
light verb. *See* verb
lingua franca 219, 438, 489, 556
linguistic-ethnographic 258, 263
Lisbon 74, 113, 118, 121, 130, 231, 232, 363
literacy 78, 102, 184, 360, 487, 488, 608, 610
loanword 119, 139, 142, 149, 205, 208–10, 393, 439, 530
locality principle 606
logic 31, 86, 88, 137, 138, 229, 402, 416, 465, 470, 472
Lusíadas 116, 570
Lusitanian 229, 231, 232–39, 556–60, 572
  re-lusitaniation 74
Luso-Brazilian 230, 233, 240, 568
Lusophobe 233
Lusophony 214, 369, 394, 566, 569, 572, 573
Lusotropicalism 236, 572
Luso-Tupi 557

M
Macau 314
Maciel, Maximino 31
mameluco 62, 64, 67, 68, 93, 494

Maranhão 64–73, 125, 261, 490–494
masculine (gender). *See* gender
Mato Grosso 73, 270, 271, 534, 539
Mattoso Câmara, Joaquim 34–36, 88, 460
Maurer, Theodoro Henrique 34–36, 460
meaning 134–49, 171, 371–81, 385, 388–95, 453–55, 459, 461, 463, 466, 469–77
media 107, 108, 319, 320, 324, 439, 440, 503–9, 553, 564, 565
medieval 117, 213, 517
melody 361, 362, 395, 543
memory 592–95, 601, 603, 609
mental 143, 245, 579, 583
MERCOSUL 572, 573
metalanguage 114, 135, 450, 497
metaphony 117, 119, 293
metaphor 81, 149, 200, 212, 380, 381, 388
metaplasm 200
metathesis 114, 130
methodology 197–99, 288–90, 296, 299, 300, 401, 445, 450, 580–82, 592, 593
metonymy 149, 380, 381, 388, 393
metrical 291, 295, 339
micro-categories (sociolinguistics). *See* sociolinguistics
micro-dialectal. *See* dialect
microstructure (lexicography) 197, 198, 446, 453
middle class. *See* social class
mid vowel. *See* vowel
migration 59, 64, 69, 71, 88, 108, 208, 209, 264, 272, 283–89, 294, 298, 301, 302, 312, 325, 442, 548
Minas Gerais 68, 73, 93, 95, 204, 263, 500, 507, 523, 534
minimalist program 588
minority language 68, 80, 206, 438
miscegenation 93, 96, 107, 220, 229, 230, 234, 535, 572
Misiones (Argentina) 311, 317–25
missionary 23–25, 70, 93, 491, 494, 519, 521, 553, 555
mixed grammar. *See* grammar (system)
mixed-race 59, 62–64, 67, 73–77, 93–96, 489, 497, 554, 561
Mobral 6, 360
modernism 256, 288, 501
modification 293, 371, 372, 376, 377, 520, 534, 544, 594, 602, 604
monolingual 80, 85, 310, 323, 326, 442, 443, 449, 452, 453, 463, 556, 562, 563, 589, 605
monophthongization 114–19, 123, 235, 499
monosyllable 338, 567
mood 171, 361, 373, 376, 377, 473, 478, 503
morph, 134, 135, 371, 374, 377, 378
morpheme 134, 135, 138, 140–50, 339, 349, 359, 370, 374, 380, 391, 392, 396, 397, 461, 514, 587, 588
morphology 133–50, 170, 223, 224, 291–96, 369, 370, 373–82, 385–97, 584–89, 591, 609
morpho-phonetic 403
morphophonological 134, 135, 145, 295, 588, 590, 591, 603
morpho-syntax 135, 137, 145, 146, 156, 157, 169, 171, 197, 201, 220, 226, 228, 266, 271, 422, 424, 426, 590
movement (syntax) 179, 180, 224, 383, 417, 429–31
Mozambique 215, 569, 571
multi-dimensional (atlas). *See* atlas
multilingualism 86, 87, 92, 93, 107, 494, 556, 580
multivocabulary unit 452, 453

N

narratio 502
narrative text 489, 496, 542, 570
nasal 116, 119, 120, 124–30, 140, 301, 312, 313, 318, 346–51, 360, 383
  floating n. 349, 350
nasalization 116, 128, 260, 295, 337, 346–49, 360
Nascentes, Antenor 33, 120, 124, 203, 233, 237, 252–57, 272
nation 221, 230–33, 236–39, 557–72
nationalism 28, 32, 221, 229, 231, 234, 236, 239, 252, 256, 277
nations (native) 54, 55, 67, 543
native peoples/languages 24–28, 56, 61, 70, 73–77, 92, 201, 252, 438, 448, 470, 478, 489, 514–16, 518–26, 531–43, 563
nativization 86–89, 92–94, 97, 107
near-bilingualism. *See* bilingualism
negation 137, 163, 285, 352, 381, 424
neo-grammarian 290

neologism 211, 213, 393, 437, 441, 445, 451, 519, 558
Neto, Serafim da Silva 30, 33, 34, 206, 238, 253, 556
neuroscience 15, 580, 581, 608, 609
neuter. *See* gender
neutralization 120, 312, 313, 341–44, 383, 388, 406, 409
New Holland 522
newspapers 165–68, 179, 184, 186, 252, 502–9
Nheengatu 24, 27, 68, 81, 93, 490, 542, 555
Niger-Congo 59, 204
Nigeria 59, 70, 530
Nilo-Saharan 59
nomenclature 443, 444, 449
nominal 45, 88–90, 165, 372, 590
nominative 104, 157–61, 226, 235, 403, 407, 412
non-autonomy 401
non-compositional. *See* compositionality
nonconcatenative 369, 370, 391, 394–97
non-dependent 164
non-deterministic 44
non-discrete categorization 380, 401
non-final (segment). *See* segment (phonology)
non-finite 171
non-homophonous 569
non-palatalization 312
non-productive 143
non-prototypical. *See* prototype theory
non-referential. *See* reference
non-salient. *See* salience
non-standard. *See* standard
non-verbal 377, 441
non-verb. *See* verb
non-white 63, 72, 74, 77
norm 14, 36, 37, 40–44, 53, 74, 137, 156, 164, 219–22, 239, 240, 358, 487, 501, 503, 560, 563
noun 144–48, 227, 228, 293, 294, 373, 408, 474–78, 534, 584–87
  bare nominal 11, 105, 227, 228, 477–79
  pseudo-n. 585
nuclear stress. *See* stress
nucleus (syllable). *See* syllable
null object. *See* object
null subject. *See* subject
number 98–105, 136, 170, 171, 375, 376, 584, 587, 588, 609
NURC 37–39, 297, 298, 358, 463

O

object 104, 106, 119, 160, 162, 165, 166, 181–84, 226, 228, 244, 285, 295, 296, 352, 412, 418, 421, 430, 505, 563, 593–99, 608
  null o. 161, 164–66, 181, 186, 227, 229, 243–45, 412, 462, 594–98
obligatority 162, 185, 222, 224, 228, 370, 374, 421
oblique 159, 161, 285, 407
occlusive 203, 311, 338
Old Portuguese 90, 113–16, 122, 127, 128, 138–40, 145, 146, 150, 156–61, 169–75, 212, 213, 232, 403–6, 410, 461, 475, 560
omission 320, 410–12, 419, 420, 461, 477, 534
one-dimensional (atlas). *See* atlas
onomasiological 204, 263
onomastics 513, 515, 539, 540, 544, 547, 548
onset (syllable). *See* syllable
onymization 513, 534
open syllable. *See* syllable
open vowel. *See* vowel
open word class. *See* word class
operator 472, 476
optional 144, 185, 347, 370, 420
oral language 215, 405, 427, 488, 489, 494–500, 503, 508, 582, 584
oral vowel. *See* vowel
orixá 438, 530
orthographic unification 553, 566–70
orthography 215, 440, 553, 566–75
  orthographic accent 117, 569, 570
  simplified o. 567
over-differentiation 313
overt expression 179, 225, 294, 421, 594–98
Ovibundo 205

P

palatal 118, 129, 130, 300, 320, 348, 351, 353, 357–60
palatalization 117, 123–25, 128, 130, 318, 320, 337, 351–53, 357, 358
  regressive p. 352, 353
Papiamento 94
Pará 67, 68, 73, 125, 272, 353, 526, 527, 529, 536
paragogic 119
Paraguay 317, 318, 325
Paraíba 63, 260, 353, 354, 358, 529
parameter 11, 240, 241, 402, 589

null subject p. 167, 187, 296, 406, 410–14, 419
Parametric Sociolinguistics 290–92, 296, 297
Paraná 121, 266–69, 311–13, 346, 519, 527
parataxis 425
parole 134, 149, 219
paroxytone 339, 340, 570
parsing (syntactic) 583, 598, 604, 606
participle 139, 162, 372, 373, 420, 424, 461, 473, 475, 476
particle 99, 105, 362, 372, 377, 385, 392, 393, 431
parts of speech 25, 41, 472
passive 243, 420, 421, 473, 474
analytical p. 476
paulista 264, 265, 269
pejorative 148, 149, 372
perception 28, 230, 284, 293, 295, 301, 302, 363, 590, 591
perfect 117, 126, 378, 465, 503
perfective 376
performance 90, 291, 588
periphery 94, 300–302, 503
periphrasis 171–74, 187, 228, 420, 424, 425, 465, 473, 476, 477
person 90, 100, 101, 104, 105, 138, 167, 170–72, 181, 224–27, 294, 321, 325, 374–77, 403–7, 590
PEUL 6, 7, 37, 298
philology 23, 28, 32, 36, 40, 232, 235
philosophy of language 467, 470
phoneme 126, 129, 144, 145, 148, 149, 203, 272, 318, 320, 337, 526, 529
phonetic 90, 114–18, 267–75, 278, 348, 361, 383, 567
phonic salience. *See* salience
phonology 113–16, 125, 126, 235, 290–95, 337, 338, 348–52, 359, 383, 384, 387, 391, 396, 567, 590, 591, 607
phonotactics 119
PHPB 8, 42, 43, 173, 184, 196
PHPP 8
phrase 361, 401, 414–17, 423
phraseology 444, 452–55
Piauí 73, 114, 261
pidgin 64, 68, 89, 92, 96–98
pied-piping 183, 184
pitch 361–64
plantation societies 94–99
Plaut-dietsch 311
plosive 114–20, 125, 128–30, 149, 337, 338, 347, 348, 351–53
pluperfect 376
plural 87, 146, 157, 170, 234, 235, 238, 292–94, 312, 320–22, 359, 384, 403–5, 466, 471, 587, 588
plurilingual 286
polarization (sociolinguistic) 79, 86, 99, 105–8
policy (linguistic) 70, 72, 81, 94, 239, 460, 494, 520, 553–57, 564, 566, 575
Polish 310–13
politeness 157, 158, 224, 381
politics 72–75, 219, 230, 237, 240, 502, 520, 531–33, 558, 565, 566, 570–76
polysemy 140, 141, 469
Pombal 69–72, 78, 494, 520, 555, 556
Pomeranian 10, 311, 541
portmanteau 142
Portugal 23–25, 74, 75, 118, 121–25, 150, 213–15, 231–40, 557–59, 563, 566, 568–75
Portunhol 319–24
possession 175–77
possessive. *See* pronoun
postpositions 25, 26
post-tonic. *See* tonicity
post-verbal 178, 179, 185, 413, 420
PP-chopping. *See* preposition
pragmatics 158, 286, 287, 378–81, 395, 397, 409, 412–14, 417–19, 421, 460, 462
pre-colonial. *See* colonialism
preconsonantal. *See* consonant
predicate 41, 105, 172, 296, 408, 409, 422, 423, 472
predication 88, 407, 409, 413, 417, 419–23
pre-dorsal-dental 358
prefix 139, 142, 352, 386–91, 441
prejudice (linguistic) 36, 37, 42, 99, 209, 234–36, 564, 565
prepalatal 114, 116, 119, 129, 320
preposition 25, 26, 106, 161, 172, 181–84, 226, 235, 285, 325, 408, 428, 581
PP-chopping 183–85
prescriptive 41, 42, 239, 288, 289, 460, 505, 531, 563
presentational verb. *See* verb
present tense 126, 172, 173, 377, 465, 476
preservation 93, 115–21, 138, 219, 240, 345, 346, 352, 383, 384, 514, 541, 559, 565

prestige 74, 107, 124, 164, 209, 278, 299, 319, 320, 352, 488, 498, 508, 560, 564
presupposition 467
pretonic. *See* tonicity
pre-verbal 162, 177, 179, 180, 185, 186, 224, 225, 419
priming 582, 583, 595, 608
principle of information distribution 416
principle of phonic salience 293
Principles and Parameters 43, 240, 291, 296, 402, 604, 606, 609
processing 469, 579–84, 590–609
proclisis 160–64, 222, 223, 228, 234
pro-drop 167, 169, 292, 296, 596, 598
  semi-pro-drop language 596, 609
productivity (morphology) 134, 139, 140, 144, 147–50, 371, 378, 379, 386, 388, 393, 395, 441, 537
progressive aspect. *See* aspect
progressive palatalization 353
PROHPOR 43, 135, 464
projection 102, 290, 361, 409, 423
prominence 124, 130, 170, 291, 340, 361, 362, 411, 421, 425, 448, 601
pronominal 157–61, 226, 402–7, 596, 601
pronoun
  demonstrative p. 102
  personal p. 104, 115, 137, 157, 166, 167, 184, 221, 244, 294, 403, 406, 409, 428
  possessive p. 157–59, 161, 175–77, 227, 228, 320, 403
  reflexive p. 97, 105, 136, 226, 244, 285, 421, 426, 476
  resumptive p. 183, 184, 285, 291, 426, 430
pronunciation 26, 113–26, 130–32, 226, 320, 353, 357, 358, 383, 390, 410
proparoxytone 260, 339, 340
prosody 291, 361, 363, 387–90, 395, 396, 409, 606
  fall-rise p. 363
prototype theory 379, 380, 387, 388, 461, 469
  non-prototypical 148
pseudo-cleft. *See* cleft
pseudo-determiner. *See* determiner
pseudo-etymological. *See* etymology
pseudo-noun. *See* noun
pseudosuffix. *See* suffix
pseudoword. *See* word
psycholinguistics 98, 469, 579–88, 592, 594, 598, 605–9
psychology 80, 149, 241, 374, 459, 547
purism 30, 75, 220, 229–36, 501, 553, 565, 566

Q

Quantification 465, 466, 471
quantifier 320
quantitative 165, 168, 185, 224, 225, 286–89, 291, 296, 297
question 361–63, 518
quilombo 72, 96, 208, 530, 537

R

raising (vowel). *See* vowel
r-caipira 121, 122, 126, 129
real time 270, 296, 601
reanalysis 171, 177, 404, 425, 583, 605
Recife 122, 125, 126, 358, 359, 522
re-compounding. See compound
redundancy 97, 136, 141, 145, 146, 384, 534, 584, 586–88, 599, 601
reduplication 394–96
reference 144–49, 168, 177, 244, 371, 419, 505, 594–98
  non-referential 422
referent 296, 409, 411, 416, 419, 428, 469, 477, 537, 584, 592, 593, 600
reflexive 97, 105, 136, 226, 420, 476, 598, 599
regionalisms 437, 447, 448, 530, 537
register 206, 285, 405, 412, 423, 424, 427, 452, 493, 499
relative clause 163, 169, 183, 184, 285, 291, 401, 402, 425–29, 471, 581, 592–94, 602–9
religion 25, 26, 56, 61, 70, 72, 81, 94, 136, 204, 230, 438, 488, 493, 500, 508, 509, 516–18, 521, 530, 532, 543, 547
religious text 491–500, 504, 507
re-lusitaniation. *See* Lusitanian
representation 134, 142, 241, 338, 348–50, 360, 388, 454, 460, 466, 478, 590, 591, 603, 604
restriction 105, 166, 169, 171, 175, 177, 207, 224, 227, 228, 242, 244, 286, 287, 296, 364, 371, 382, 389, 409, 411, 417, 419, 424, 429, 542
restrictive clause 427, 428
restructuring (linguistic) 71, 89, 97, 98, 106, 406, 419
resumptive pronoun. *See* pronoun

resyllabification. *See* syllable
retroflex 121, 122, 269, 274, 278, 301–3, 337, 356
rhetorical interrogative. *See* interrogative
Rhineland 311
rhizomatic 288
rhotacism 126, 278, 499
rhotic 121–23, 128, 130, 351, 354–56, 358
rhythm 10, 228
Ribeiro, Júlio 31
right-branching 592, 593
Rio de Janeiro 34–37, 63, 73, 74, 76, 102, 121–25, 263, 264, 285, 289, 361–63, 495, 500, 503
Rio Grande do Norte 113, 120, 342, 523
Rio Grande do Sul 73, 120, 121, 310–12, 325, 342, 345, 346
rising (syntax) 362, 363
Romance Languages 30, 31, 40, 90, 117, 120, 137–40, 162, 171, 195, 200, 226, 227, 242, 473–77, 604, 605
Romanian 137, 527
Romanticism 30, 77, 256, 501, 531, 557, 572
Roraima 315
rural 75–79, 90–95, 99–102, 105–8, 266, 267, 277, 278, 285, 313, 317–19
rurban 79, 80, 289
Russian 34, 370, 374, 527, 531, 532, 541

S

salience 104, 120, 222, 293, 295, 405, 414, 595–98
  non-salient 595–97
Santa Catarina 64, 117, 121, 125, 161, 266, 269, 310–12, 346, 360, 534
São Paulo 9, 73, 76, 93, 263–66, 285, 286, 294–302, 312–20, 505
Saussure, Ferdinand de 34, 88, 149, 219, 537, 567
school 28, 42, 61, 71, 72, 78, 164, 208, 232, 234, 245, 261, 264, 319, 322, 324, 351, 356, 379, 443, 449, 494, 496, 500, 508, 544, 545, 554–65, 589
segment (phonology) 101, 143, 318, 337, 338, 341, 347, 348, 350, 354, 357–60, 383–85, 388, 391
  non-final s. 126, 344, 345
segmentation (morphology) 138–40, 143, 144, 377, 378
self-paced reading 582–84, 595–600, 604, 605
semantic equivalence principle 455
semantics 140, 141, 144, 145, 171, 263, 380–82, 385, 388–90, 409, 410, 459–80
semasiological 446, 453
semi-auxiliaries 473
semi-creole 88–91, 321
semi-orthographic transcription. *See* transcription
semiotics 464
semi-pro-drop language. *See* pro-drop
semivowel. *See* vowel
Senegal 59
sentence 409, 412, 423
  complex s. 416, 422–24
sex 62, 76–78, 144, 293–95, 298, 360, 370, 371, 584, 600
shortening (morphology) 130, 393–95
sibilant 114, 117, 123, 124, 352
sign 144–49, 212, 579
signified 145–49, 198, 436
signifier 145–49, 198, 200, 211, 436, 537
sign language, Brazilian 15, 81, 563
simplification (morphology) 86, 88, 99, 100, 103, 105, 238
simplified orthography. *See* orthography
singular 293, 294, 320, 321, 325, 374–79, 403, 405, 478, 587, 588, 596
slavery 59, 60, 65, 67–78, 86–100, 114, 118, 119, 166, 183, 202, 494–98, 529
Slavic 310, 313, 532
social 3, 75, 120, 206, 270, 278, 295, 405, 446, 515, 564, 580
  s. class 70, 72, 78, 79, 99, 100, 120, 124, 231, 285, 287, 292, 300–302, 322, 324, 495, 500, 502, 508, 544
  middle c. 105–8, 303, 352
  upper-c. 105
  upper-middle c. 303
social-historical 194
social strata 59, 74, 75, 270, 273, 352, 353, 364, 499, 504
Sociedade Protetora dos Desvalidos 498
sociocognitive 293, 460
sociocultural 401, 500
socio-dialects 198, 405
socio-economic 68, 75, 94, 544
socio-historical 72, 89, 92, 98, 194, 201, 208, 278, 405

socio-indexical 286, 288, 295
socio-interactional 492, 554
sociolects 197, 322
sociolinguistics 78, 79, 85–93, 97–100, 105–108, 261–63, 283–303, 358, 359
s. micro-categories 300
sociolinguistic variable 62, 197, 273, 277, 278, 285–87, 290–302
sociology 198
sociophonetic 313
sociopolitical 502, 503
solecisms 74, 77
sonorants 354
sonority 352
sonorization 200
sound atlas. *See* atlas
Spanish 60, 65, 67, 113, 114, 122, 174, 310, 313–25, 596, 597
specificational 416
specificity 165, 166, 244, 385, 417
speech act 212, 219, 487, 502
speech community 90–92, 97, 100, 174, 193, 206, 207, 241, 299, 310, 326, 353, 358, 361, 364, 375, 453
spelling 114, 115, 121, 212, 215, 348, 439, 538, 566, 567
splinter 142, 370, 387, 391–94, 397
Sranan 94
stability 92, 107, 117, 184, 187, 226, 303, 316, 323, 324, 349, 360, 401, 588
stabilization 179, 210, 324, 349, 588, 594
standard 75, 80, 92, 99–108, 113–22, 125, 214, 220, 221, 285, 297, 322, 429, 430, 441, 503, 505, 558–60, 563, 572, 586–89
non-s. 79, 183, 299, 301, 322, 429, 587, 588, 596, 609
standardization 77, 96, 126, 130, 239, 319, 323, 494, 500, 502, 508, 557, 568, 574
stigmatization 42, 53, 79, 165, 223, 295, 319, 320, 405, 428
strategy of linguistic promotion 215, 240, 572, 573, 576
stress 115, 116, 119–22, 126, 294, 321, 339–44, 347, 348, 352, 359–63, 370, 383, 384, 395
nuclear s. 362, 363
structuralism 35–38, 134, 143, 436, 460, 462
Structural Semantics 464, 466
style 80, 136, 137, 145, 221, 232, 240, 286, 287, 292, 300, 301, 352, 395, 474, 487, 504, 505, 560
subculture 507
subdialect. *See* dialect
subject 160, 166–70, 178, 180, 185, 225, 229, 243, 296, 410–14, 418, 421, 430
null s. 167, 169, 180, 410–14
subject-focused 593
subject-prominent language 11
subject-topic 180
subject-verb order 90, 221, 234
subordination 146, 147, 180, 235, 409, 413, 598
substrate 33, 98, 106, 115, 122, 204
subsystem. *See* grammar (system)
subtonic. See tonicity
suffix 119, 134, 139–44, 147–49, 339, 340, 349, 371, 373, 382–90, 539, 590
pseudosuffix 141
suppletion 159, 295, 378
suprasegmental 130
syllable 114–16, 119, 121, 128, 130, 134, 295, 312, 338–42, 347–62, 382, 591
CV-s. 130
nucleus (syllable) 130, 338, 395
onset (syllable) 311, 337, 338, 347, 350, 351–56, 382, 396
open s. 10, 338, 352, 396
resyllabification 356, 384
tautosyllabic 116
three-syllables window constraint 339
syllable-timed 10
symbol 144, 145, 148, 299, 472
synchrony 135–43, 146, 149, 460, 466
syncope 499
synecdoche 212
synonym 36, 139, 233, 234, 459, 473
syntactically-driven morphology 370
syntactically-motivated voicing. *See* voicing
syntagma 136, 147, 444, 452, 453, 526, 535, 536
syntax 146, 155–69, 178, 184–86, 226–28, 234, 235, 291, 292, 295, 370, 401–31, 460–62, 592–94, 603, 604
synthetic future 171–73
synthetic passive 476
Syrian 211

T
Tagmemics 36
Taiwanese 314
tautosyllabic. *See* syllable
television 316, 322–25, 504–6, 509
temporal semantics 171, 465, 468
tense 116, 126, 171–73, 228, 373–78, 465, 466, 471, 473, 476
tense-mood-aspect 99, 105, 369, 375, 376
terminology 197, 199, 440, 443–45, 574, 575
text 165, 173, 175, 445, 446, 469, 472, 479, 487, 488, 490–503, 508, 564
textbook 210, 462, 507, 565
thematic role 410, 420, 421
thematic verb. *See* verb
thematic vowel. *See* vowel
theory 28, 30, 36, 39, 198, 241, 287–89, 464, 468, 478
three-syllables window constraint. *See* syllable
Tirolean 312
tonicity 342, 348, 352, 361, 362, 567
  post-tonic 126, 127, 344–46, 352, 362
  pretonic 114, 117–19, 126, 127, 254, 260, 261, 273, 278, 342–44, 347, 352, 362, 383, 388
  subtonic 569
top-down change 80, 508, 509
topic 11, 168, 169, 178–80, 229, 244, 409, 413, 414, 417–20, 596
topicalization 412, 414, 417–19
topic-oriented 12, 169, 179, 181, 224
topodynamic atlas. *See* atlas
toponomastics 513, 514, 548
toponymy 203, 442, 513–40, 548
topostatic atlas. See atlas
trace 180, 409, 421, 424
traditional grammar 38–42, 75, 147, 149, 288, 393, 441, 475, 508
transcription 6, 8, 24, 38, 259, 278
  semi-orthographic t. 38
transfer 98, 106, 205, 207, 210, 513, 538, 543
transitive 178, 413, 477
translation 442, 453–55, 472, 518
Trentino 312
truth-conditions 471, 472, 478
Tupi 23–29, 122, 201–4, 489–92, 515, 517–29
Tupi-guarani 54, 55, 93, 489, 515
Tupinambá 23–26, 56, 67, 68, 93, 201, 489, 490, 541, 542
Turkish 134, 211
two-dimensional atlas. *See* atlas
Tycho Brahe Corpus 8, 9, 168, 228, 464
typology 1, 10–12, 85, 86, 369, 370, 394

U
Ukrainian 309, 310, 313, 531, 532
Umbundu 65, 204, 205, 530
umlaut 569, 570
unacceptable 598–601
unaccusative 179, 224, 229
unanimated 296
underlying structure 241, 243, 349, 353
underspecified 584
ungrammaticality 106, 174, 180, 375, 582, 587, 604
uniqueness 461, 540, 545
unity of the Portuguese language 2, 28, 30, 193, 214, 237, 239, 245, 562
universal 98, 241, 245, 287, 290, 350, 418, 419, 598, 603–606
universes of discourse. *See* discourse
university 35–39, 42–45, 288, 579, 609, 614
univocity 444
unstressed vowel. *See* vowel
upper-class. *See* class
upper-middle class. *See* class
upper-mid vowel. *See* vowel
urban 76–79, 86, 106–8, 120, 223, 224, 285, 286, 289, 295–298, 317–20, 500–504, 532, 533
urbanization 74, 75, 78, 88, 96, 108, 261, 277, 286, 500, 509
urbanonym 522, 530
Uruguay 309, 316–25
usage of language 31, 32, 37, 39–42, 157–61, 164–77, 198, 228, 293–96, 324, 406, 436, 461, 475, 488, 493, 508, 597, 613
utterance 177, 241–45, 287, 361–64, 379, 580, 584–86

V
vagueness 147, 385
variable (sociolinguistics). *See* sociolinguistics
variation 89–92, 98–107, 116–19, 160–63, 173–76, 283–93, 299–303, 403–5
Vasconcelos, José Leite de 31, 234, 235, 238
velar 121, 122, 129, 274, 312, 320, 348, 360, 383

velarization 120–23, 128, 130, 312, 317, 337, 357
Veneto 309, 312, 517
Venezuela 67, 318
verb 170–80, 372–74, 396, 402–13, 416–25, 610
  light v. 422, 423
  non-v. 339, 340
  presentational v. 413
  thematic v. 162, 222
vernacular 37, 86, 89, 114, 164, 221, 232, 254, 290, 299, 312, 318–21, 406, 412, 441, 488, 497
vibrant 122, 129
Visconde de Pedra Branca 28, 121, 124, 206, 231, 251–53
VOC 215, 439, 440, 570
vocabulary 26, 195–98, 205–7, 230, 251, 252, 256, 265, 440, 491, 492, 497, 575
vocalic alternation 34
vocalism 115, 117, 130, 341, 342, 347
vocalization 115, 120, 127, 130, 317, 318, 320, 337, 356, 357
vocative 534
voice (verb) 372, 412, 414, 419, 420, 459, 473
voiceless 200, 311, 357
voice onset timing 311
voicing 114, 116, 122, 123, 128, 129, 200, 203, 290, 311, 318, 358
  syntactically-motivated v. 318
vowel
  mid v. 120, 312, 341–47, 350, 351, 354, 383, 388, 390, 394, 527
  open v. 114, 118, 119, 122, 126, 261, 273, 342–45, 567
  oral v. 38, 80, 126–28, 347, 498, 502, 507
  raising (vowel) 117, 119, 127, 312, 344–47, 351, 357
  semivowel 120, 121, 349
  thematic v. 138, 139, 144, 322, 349, 376–78, 382, 384, 389, 395
  unstressed v. 115, 118, 127, 130, 268, 341, 344, 347, 348, 359–61, 375, 382, 569
  upper-mid v. 383, 390, 395

W

well-formedness 241, 243, 290, 582
Westfalian 311
WH-phrase 155, 169, 185, 291, 362, 413
word
  complex w. 378, 384–89, 394, 445, 452, 453
  open word class 137, 138, 370, 387, 424, 436
  pseudoword 7
  word class 136, 138, 143, 146–49, 361, 370–77, 380, 385, 386, 390–93
  word-external 355
  word-final 90, 337, 345, 346, 349, 354–60, 395
  word-formation 369, 394
  word-initial 116, 311, 352, 353, 355
word-internal coda. *See* coda
writers (literary) 28–30, 77, 193, 205, 236, 266, 500, 501, 558, 559, 561
writing 53, 63, 64, 73, 487, 499, 507, 514
written language 150, 156–67, 173, 176, 179–86, 198, 245, 487–503, 507, 508

Y

Yoruba 70, 204, 205

Z

zero 165, 344, 374, 409